Hox genes from nematodes and RNAi in *Brugia malayi*

A. Aziz Aboobaker

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Edinburgh

2002

Declaration

I declare that this thesis has been composed by myself and, except where otherwise stated, is entirely my own work.

A. Aziz Aboobaker

Acknowledgements

I owe everything to the love and devotion of my whole extended family and this collection of thoughts and explorations symbolises the efforts of all of us. To my mother Rahat, the hard sacrifices you have made so that I could have every opportunity to succeed are the foundations for my achievements. My grandparents Zuleikha and Yousuf took it upon themselves to raise a fifth child and a third son. No words would suffice to describe the importance of your gifts to me; a safe home, a sense of pride in my identity, respect for myself and others and the self-confidence to pursue my dreams. I hope you are proud of what I have achieved. To my aunt Nigar, as a child I never felt more loved and cherished then when I was sat in your lap and I still feel the same every time I see you. To my darling little brother and sister Zain and Aisha always chase your dreams and be nice to mummy, especially when she wants to relax! To my sisters Hannah and Lauren thanks for accepting me into your lives so seamlessly and helping me maintain perspective (yes, you know what I mean).

The debt of gratitude I owe to Rajeeve for his unconditional friendship, support at all times and for inviting me into his wonderful family cannot be repaid in this lifetime. It's a long way to if you don't know where you're going and you have helped me find the way. Prima, without you Edinburgh would have been impossibly dull and without your encouragement I may have given up before I really started. Sometimes I am glad I got that nasty flu' in the winter of 97/98 that prompted you to bring me some medicine. Emma, I have never had more fun and laughter then when alone in your company, thank you for your endless love and support. I look forward to seeing the all the amazing things your talents will uncover. George, Jason and Nas (in good company!) my "real deal male" friends, thank you for making Edinburgh a fun place to be, for all those nights dancing (never enough), for keeping my feet on the ground and lifting me up when I was down. Annemieke your friendship has often been the breath of fresh air I have needed both then and now (even more so!). I have met a great bunch of people in the last year to fill the gulf left by others leaving. Between them, especially you Ben, they have made staying on for another year a fun prospect. To my new acquaintance Kate in Beanscene (whose name I now know!) thank you for big coffees and even bigger smiles. To my other best friends from my times during and before Edinburgh Nick, Kambiz, Jon, Shan, Nolwenn, Jo, Desiree, Stephanie and Co seeing you all happy always cheers me up no end, although I don't see you often enough for my liking. "Michsala" thanks for putting up with me during 6 months of "thesis hell" and making it inseparable from easily some of the best times of my life. You helped me over the finishing line and still remain the dearest of friends at the end of it all. Timing is so crucial for some things in life.

A final thank you to my co-workers that have made some of the days here as fun as some of the best nights. You are thanked at the appropriate places in the thesis for your scientific input but Mark B (Big Boss Naboo), Jen, David, Robin, Claire, Marian, Mark D, Mark W, Eyualem, Peter, Bill, David, JP, Denise, Allens, Arnots, Maizelelettes and Barton Slatko and Co at New England Biolabs thank you again.

ABSTRACT

Hox genes are important: playing a central role in the anteroposterior patterning of bodyplans, showing conservation of relative expression pattern and chromosomal order of paralogous groups between phyla, providing the framework for a molecular map of animal body plan evolution and gross morphology. One organism for which this paradigm appears to be inaccurate is in the nematode *C. elegans*, which has a depauperate Hox cluster compared to other protostomes. The data presented here reveal that Hox genes within the phylum Nematoda are undergoing rapid and dynamic evolution. Hox genes orthologous to those in other protostome phyla have been identified from species representative of a cross-section of the whole of Nematoda, which are definitively absent from *C. elegans*. This demonstrates loss of multiple Hox genes within the nematode lineage and challenges previously defined Hox signatures for the two major protostome lineages. In addition we observe alternate cis-splicing of the same N-terminal exon to two different C-terminal homeodomains from different Hox orthology groups within the parasitic nematode *B. malayi*. These findings demonstrate Hox gene loss as an alternative mechanism for the evolution of body plans to gene duplication and changes in coding and regulatory sequences, within the Nematoda. Furthermore the conservation of the simple nematode body plan has not been accompanied by coordinate Hox gene loss through the whole phylum.

One problem when working with parasitic species is the lack of molecular genetic techniques available to study gene function directly. A method for studying gene function using the conserved mechanism of RNAi is described and evaluated. The technique successfully results in loss of specific mRNA transcripts from the target organism *B. malayi*, and allows the observation of resulting phenotypes in culture. This new approach offers the possibility of investigating the roles of developmental genes such as the Hox gene family, or screening for potential novel drug target candidates and should also be applicable to other parasitic helminths.

CHAPTER 1: HOX GENES, ANIMAL PHYLOGENY AND THE EVOLUTION OF METAZOAN BODY PLANS: THE STRANGE CASE OF THE *C. ELEGANS* HOX CLUSTER

1.1. Hox genes: a paradigm for the evolution of gene function in a developmental context

Hox genes are important: for over two decades research into the nature of these fascinating transcription factors has been at the centre of our understanding of developmental biology (Akam, 1998c; Lewis, 1978; Lewis, 1998; Ruddle et al., 1994a). More recently, comparative study of these genes across phyla (Akam, 1995; de Rosa et al., 1999; Duboule and Dolle, 1989; Grenier et al., 1997), within phyla (Akam, 1998a; Cook et al., 2001) and between closely related species (Averof and Patel, 1997; Eizinger and Sommer, 1997a; Sommer et al., 1994; Stern, 1998a) has been the basis for developing our theories for the evolution of developmental mechanisms (Akam, 2000; Sommer, 1997a; Sommer, 1997b; Sommer, 1999). More generally mechanisms for the evolution of gene function and the apparently highly conserved genetic networks within which they act are starting to be understood. As more and more examples of the evolution of gene function, particularly in a developmental context, are described in the often exquisite detail afforded by cutting edge molecular genetic analysis (Grandien and Sommer, 2001; Greer et al., 2000; Grenier and Carroll, 2000) we are able to continually refine and improve our theories of evolution at the molecular level. The incorporation of these comparative analyses of individual examples with wider phylogenetic sampling (de Rosa et al., 1999), the spread of genomics to previously remote regions of the tree of life (Pollock et al., 2000) and the first signs of synergy between functional and population genetics (Delattre and Felix, 2001; Purugganan and Suddith, 1999; True and Haag, 2001) promises much. As these previously separate undertakings converge they are shaping a theoretical framework for understanding evolutionary developmental biology and the paradox of molecular conservation versus morphological diversity. At the centre of this the Hox gene cluster is still the substrate of choice for adding to our

understanding. Hox genes and the conserved 60 amino acid DNA binding domains they all encode were the first astounding example of molecular conservation across phyla (Gehring et al., 1994). Currently Hox genes are also at the centre of explaining how despite molecular conservation morphological diversity is generated (Arthur, 2000; Grandien and Sommer, 2001; Sommer, 2001; Stern, 1998b).

1.2. An outline of Hox gene evolution and function

The identification of various gene families highly conserved across metazoan taxa has provided many new insights into the evolution of animal body plans (Davis and Patel, 1999; Takahashi et al., 2001). Perhaps the best characterised of these families is the Hox gene family, which plays a central role in patterning positional identities along the antero-posterior axis (Gellon and McGinnis, 1998). Hox genes are best characterised in vertebrates and insects but are being described in an ever increasing number of phyla (de Rosa et al., 1999; Dick and Buss, 1994a; Grenier et al., 1997; Ishii et al., 1999; Kenyon et al., 1997; Mito and Endo, 1997; Mito and Endo, 2000; Peterson et al., 2000).

Hox genes are found clustered in the genomes of Bilaterian animals with multiple Hox members arranged along contiguous regions of the genome (Finnerty and Martindale, 1998). In the vertebrate lineage the whole gene cluster has undergone duplication to produce multiple clusters (Finnerty and Martindale, 1998). Furthermore Hox genes display the general characteristic of spatial colinearity (Mann, 1997; Wada et al., 1999), with their expression domains along the anterior posterior body axis corresponding to their relative 3' to 5' chromosomal order. The Hox cluster is assumed to have arisen by tandem duplication of an ancestral gene. Based on the known distribution of Hox clusters, it is clear that the origin of the Hox cluster predated the radiation of the Bilateria (de Rosa et al., 1999). In all Bilaterians surveyed Hox genes corresponding to anterior, central and posterior genes expressed in the head, central and tail regions respectively have been identified (de Rosa et al., 1999; Finnerty and Martindale, 1998). As more phyla have been sampled it appears that the Hox orthology groups (a group of individual genes from different species that that were represented by a single gene in the common ancestor) present in extant

Bilaterians are shared and were therefore present in the common ancestor of all Bilaterians (de Rosa et al., 1999). However this interpretation of Hox gene evolution has only been made after the recent radical reorganisation of metazoan phylogeny (see below) and an expansion in the number of phyla sampled for Hox genes.

In addition to the existence of an expanded Hox gene cluster at the root of the Bilateria (de Rosa et al., 1999), studies of in the cephalochordate Amphioxus have identified an ancient paralogous gene cluster (Brooke et al., 1998). This cluster of genes appears to be an ancient evolutionary paralog of the Hox cluster consisting of three genes called the ParaHox cluster (Brooke et al., 1998). In Amphioxus the ParaHox cluster consists of three genes, *AmphiXlox* a member of the *Xlox* family, *AmphiGsx* a member of the *Gsx* family and *AmphiCdx* a member of the *caudal* or *Cdx* gene family. Phylogenetic comparison between these genes (also present in protostome phyla) and Hox cluster genes indicate the *Cdx* family is an ancient paralog of posterior Hox genes, the *Xlox* gene is related to *Hox3* group genes and the *Gsx* family is most closely related to anterior Hox genes. In addition this cluster of genes obeys the principle of spatial colinearity (Brooke et al., 1998; Holland, 2001). It appears that at some point in Hox gene evolution a ProtoHox cluster duplicated to produce the Hox cluster and the ParaHox cluster (Brooke et al., 1998; Finnerty and Martindale, 1998). Phylogenetic analysis has since attempted to use this information to recreate the pattern and order of duplication events that led to the ancestral Bilaterian Hox and ParaHox clusters (Ferrier and Holland, 2001; Finnerty and Martindale, 1999; Kourakis and Martindale, 2000; Schierwater and Desalle, 2001).

All Hox genes encode a conserved 60 amino acid DNA binding motif known as the homeodomain (Gehring et al., 1994; McGinnis et al., 1984a; McGinnis et al., 1984b). By interacting with known and as yet uncharacterised cofactors Hox genes regulate the expression of extensive sets of downstream target genes during development (Palopoli and Patel, 1998; Roch and Akam, 2000; Scott, 1999; Wiellette and McGinnis, 1999). Although Hox proteins regulate transcription with a high degree of biological specificity, they exhibit very similar DNA binding specificities i*n vitro* (ref). Specificity is achieved by Hox genes binding the regulatory regions of downstream target genes by cooperative binding with cofactors, such as members the PBX family of homeodomain proteins (Chan et al.,

1997; Chang et al., 1996). The Hox gene interaction with PBX family genes is mediated by the conserved hexapeptide motif that lies to the N-terminus of the homeodomain (Passner et al., 1999; Piper et al., 1999b). The hexapeptide motif reaches out and binds into a hydrophobic pocket on the surface of the PBX homeodomain (Passner et al., 1999; Piper et al., 1999b). The homeodomains of these two protein families bind DNA in a cooperative manner on opposite faces of the double helix. This interaction is believed to be important in providing Hox gene specificity for downstream targets (Passner et al., 1999; Piper et al., 1999b). However, this interaction is likely to represent only a basic unit for the control of transcriptional activity and many other cofactors are likely to be involved in the function of different Hox genes to add to temporal and positional specificity (Abramovich et al., 2000; Asahara et al., 1999).

In arthropods and chordates Hox genes are expressed embryonically and are required for embryonic development to proceed correctly (Akam, 1989; Krumlauf, 1994). In the nematode *C. elegans* half of the Hox genes are not required for correct embryogenesis (Kenyon et al., 1997), but are employed in larval maturation and the specification of cell identities along the anteroposterior axis (Kenyon et al., 1997). Although functional data from other phyla is not available yet, expression patterns suggest that in sea urchins (Arenas-Mena et al., 1998) and in some protosome lineages (Giusti et al., 2000; Kourakis et al., 1997) Hox genes may not be required for embryogenesis but for development of the adult body plan. From these data are not clear whether the embryonic requirement for Hox genes observed in vertebrates and arthropods is ancestral or derived. Some authors suggest that, the development of free-living characteristically bilaterian larvae without the need for expression of most Hox cluster genes during embryogenesis, means that Hox genes were not ancestrally required for embryogenesis in the Bilateria (Peterson et al., 1997). However, the ancestral state at other parts of extant Bilaterian evolution may have been a requirement for Hox genes during embryogenesis. Only expression and functional analysis across a broader sample of phyla will help answer this point.

1.3. Tracing Hox gene evolution over the new Animal Phylogeny

Over the last four years the shape of animal phylogenetics among the Bilaterians has undergone radical changes (Adoutte et al., 2000; Aguinaldo et al., 1997; Halanych et al., 1995; Peterson and Eernisse, 2001; Ruiz-Trillo et al., 1999). This has been catalysed by a refinement in molecular phylogenetic methods and explosion in the amount of data available for analysis (Peterson and Eernisse, 2001). The profound differences between the old and new pictures of animal phylogeny have led to a reappraisal of all the available data about the evolution of developmental mechanisms within the new framework (Adoutte et al., 2000). This is particularly true for Hox genes, broader sampling of which has been interpreted as providing independent molecular evidence for the new phylogeny (de Rosa et al., 1999).

1.3.1. The New Animal Phylogeny

Previous to this recent reappraisal of animal phylogeny many proposed evolutionary scenarios were based on traditional morphological analysis. These analyses were based on the comparison of morphological characteristics with shared characteristics used to support the alliance of different groups within the animal kingdom. These analyses led to a widely accepted view of animal phylogeny splitting the Bilateria into the Protostomia (consisting of annelids, arthropods and molluscs) and the Deuterostomata (including chordates and echinoderms) with the pseudocoelomate Nematoda and acoelomate Platyhelminthes as basal Bilaterians (Fig 1.1). Although alternative morphological analyses grouped the Nematoda with the protostomes as opposed to occupying a basal position (Fig 1.1, Nielsen). Early molecular evolutionary frameworks agreed with the basal position of the Nematoda, significantly suggesting that *D. melanogaster* and humans were more closely related than either was to the model nematode *C. elegans* (Sidow and Thomas, 1994). This view was also supported by the available data from developmental biology which portrayed the cell-lineage mode of development of *C. elegans* as simplistic and basal to the more complex modes of other Bilateria that gave rise to (allegedly) more

complex body plans (Sulston et al., 1983). The apparently primitive structure of the *C. elegans* Hox cluster also agreed with this view as it had fewer genes that nonetheless were representative of the anterior, central and posterior genes of arthropods and vertebrates (Burglin and Ruvkun, 1993; Burglin et al., 1991; Kenyon, 1994).

More recent molecular phylogenetic analysis employing mainly the 18S rDNA (18S) gene has completely restructured this view (Aguinaldo et al., 1997). Previous analysis using this gene and other molecular loci suffered from the fact that the data available was phylogentically restricted to specific taxa, some of which had fast rates of evolution. This was particularly true for the Nematoda represented by *C. elegans*, leading to them being placed in an artificially basal position. Aguinaldo et al reanalysed 18S sequences with more slowly evolving taxa splitting the Protostomes into two major clades (Fig 1.1). These are a clade of moulting animals, the Ecdysozoa (including nematodes, arthropods, tardigrades, onycophorans, and priapulids), and a second clade comprised of spirally cleaving animals and lophophore-bearing animals (including molluscs, annelids, nemerteans, bryozoans and brachiopods). Other studies using 18S on sections of animal phylogeny have agreed with this finding and analysis of Hox genes (see below) in this framework has provided independent support (Balavoine, 1997; de Rosa et al., 1999). Another approach has identified a specific molecular marker (characterised by horseradish peroxidase immunoreactivity) that appears to be specific to the nervous system of all Ecdysozoan phyla tested (Haase et al., 2001). One complication of this view is the proposed polyphyly of the platyhelminthes, suggesting that one group of this phylum (the Acoela) is in fact a sister group to the Bilateria (Ruiz-Trillo et al., 1999).

This new view of animal phylogeny posed the problem that previous painstaking morphological analysis was incongruent with molecular analysis. However, a recent combined reanalysis of morphological and molecular data (including Hox genes presence/absence as morphological characters), that dwarfs previous analyses, finds the incongruency is not very great (Peterson and Eernisse, 2001). In particular the rigorous reanalysis of morphological data is largely in agreement with 18S sequence data.

Fig 1.1 Three different phylogenetic trees for the Bilateria

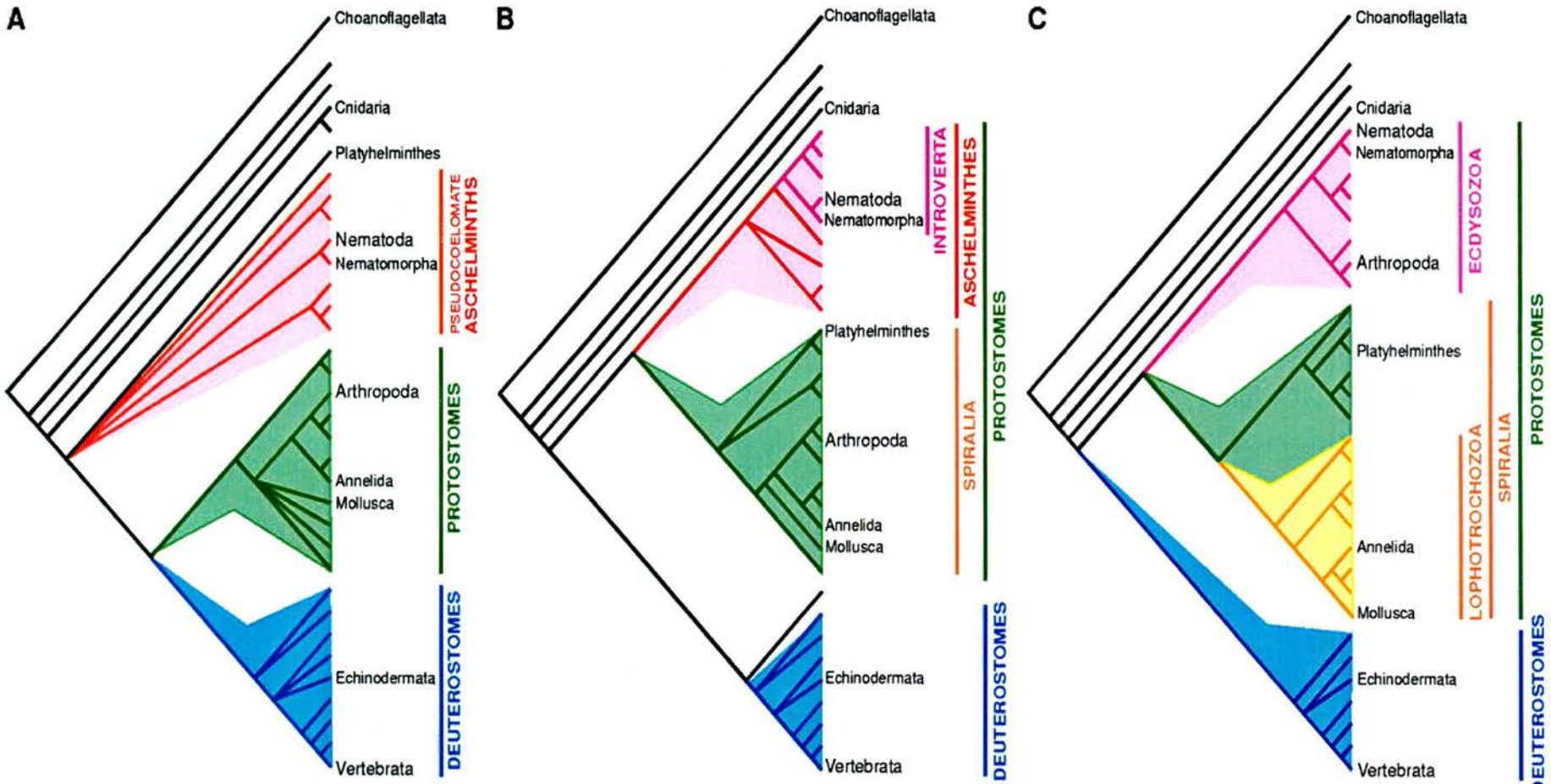


Fig 1.1 Three different phylogenetic trees for the Bilateria

Three different hypotheses for the relationships of the animal phyla. Three hypotheses of these relationships are represented; each has different implications for the expected similarity of different Bilaterian phyla. **(A)** A phylogeny based on traditional morphological criteria (Brusca and Brusca, 1990). **(B)** The phylogeny proposed by Nielsen, wherein nematodes are considered as protostomes and are grouped with other phyla having an anterior introvert organ (Nielsen 1995). **(C)** A molecular phylogeny based on that proposed by Aguinaldo et al 1997, joining arthropods and nematodes in a clade of moulting animals and other protostomes into the Lophotrochozoa, supported by further evidence from Hox gene sequence analysis (de Rosa et al., 1999) and a reconsideration of available morphological data (Peterson and Eernisse, 2001). Figure adapted from Blaxter, 1998.

1.3.2. Overlaying Hox gene evolution on the new phylogeny

Hox genes are being cloned and studied from an ever growing number of taxa and the information they provide as to the evolution of animal body plans has changed earlier views about the extent of molecular conservation (de Rosa et al., 1999; Finnerty and Martindale, 1998; Holland, 1999). While all Bilateria are likely to have Hox gene clusters they are not as conserved as previously thought. Instead it appears that the ancestral Hox cluster has undergone many secondary modifications including cluster duplication (Amores et al., 1998; Holland and Garcia-Fernandez, 1996), the acquisition of new genes by lineage specific duplications (which could also be interpreted as loss in other lineages) (Kourakis and Martindale, 2001) and unambiguous gene loss in some lineages (Aparicio et al., 1997). In addition some Hox gene sequences have evolved so that they have signature peptides within and flanking the conserved homeodomain region (de Rosa et al., 1999; Telford, 2000c). In some cases these signatures provide strong independent evidence for the grouping of the Protostomes into the Lophotrochozoa and Ecdysozoa.

The current minimum estimate for the number of Hox genes in the common ancestor of Bilateria is seven (Fig 1.2). These genes would have been the labial/*Hox1*, *proboscipaedia/Hox2*, *Hox3*, *Deformed/Hox4*, *Sex-combs-reduced/Hox5*, one additional central gene and one posterior gene. However, it is possible that the number of central class genes was as many as four if the central genes in different Bilaterian clades are directly orthologous and not the result of independent gene duplications (Telford, 2000c), with some orthology relationships obscured (de Rosa et al., 1999). Phylogenetic and signature peptide analysis of the central Hox genes suggests that genes of the two major protostome clades are likely to be orthologous and did not arise by independent duplications . These analyses equate the *fushi tarazu* (*ftz*) gene of the Ecdysozoa (arthropods, onycophorans and tardigrades) with the *Lox5* genes of the Lophotrocozoa (de Rosa et al., 1999; Telford, 2000a) and the *Antennapaedia* (*Antp*) with a central Lophotrocozoan gene characterised by *Lingula anatina HB1* and *Lineus sangunieus Hox7* (from here on referred to as just the *Antp* group) (Fig 1.2 and Fig 2.9 in chapter 2) (Telford, 2000a). Two further central genes present in both protostome clades (*Ultrabithorax* (*Ubx*)

and *abdominalA* (*abdA*) in the Ecdysozoa, *Lox*2 and *Lox4* in the Lophotrochozoa) appear to have arisen from a single ancestral gene and are characteristic to each clade (so far) (de Rosa et al., 1999). It remains unclear whether these genes arose by independent duplications in each lineage or whether they arose before the separation of the two protostome lineages (de Rosa et al., 1999; Telford, 2000c). The most parsimonious scenarios support a duplication event before separation of the Lophotrochozoa and Ecdysozoa (Telford, 2000c). This is the scenario that is assumed here and *Lox2* is equated to *Ubx* and *Lox4* to *abdA*. The equating of the protostome central genes in this way takes the minimum number of Hox genes in the basal protostome to 10.

However, the existence of three posterior gene in the Nematoda (and possibly two in the Priapulida) of the Ecdysozoa and of two distinct posterior genes characteristic of the Lophotrochozoa (*Post1* and *Post2*) takes the number up to as many as 11 or 12 genes (Fig 1.2), if this state is ancestral (de Rosa et al., 1999). The posterior *Abdominal B* (*AbdB*) gene of the Ecdysozoa is also characteristic of this clade, present in all of the proposed members. It is the only Hox characteristic placing the Nematoda in this group (de Rosa et al., 1999).

The number of ancestral genes in the basal Bilaterian is more difficult to estimate, as orthology assignment of the deuterostome central genes to those of the Protostomes is not easy (de Rosa et al., 1999). If orthology has been obscured by subsequent sequence evolution spatial colinearity would suggest that the *Hox6*, *Hox7* and *Hox8* genes are related to the *ftz/Lox5*, *Antp* and the *Ubx/Lox2/Lox4/abdA* ancestral gene respectively. One report tentatively suggests phylogenetic support for a *ftz/Lox5/Hox6* grouping (Telford, 2000a). In addition to this the expansion of the posterior Hox genes observed in deuterostomes to at least three genes in echinoderms, five genes in vertebrates and six in *Amphioxus* could be ancestral (Ferrier et al., 2000). This raises the upper limit of Hox genes in the ancestor of Bilateria to as many as 14, the number found in *Amphioxus* which has the largest cluster so far found (Ferrier et al., 2000). Whatever the number of Hox genes in the common ancestor of Bilateria, it is clear that the main expansion and evolution of paralogy groupings occurred before the radiation of the Bilateria (Fig 1.2).

Fig 1.2 Tracing the evolution of the Hox cluster through the Bilateria

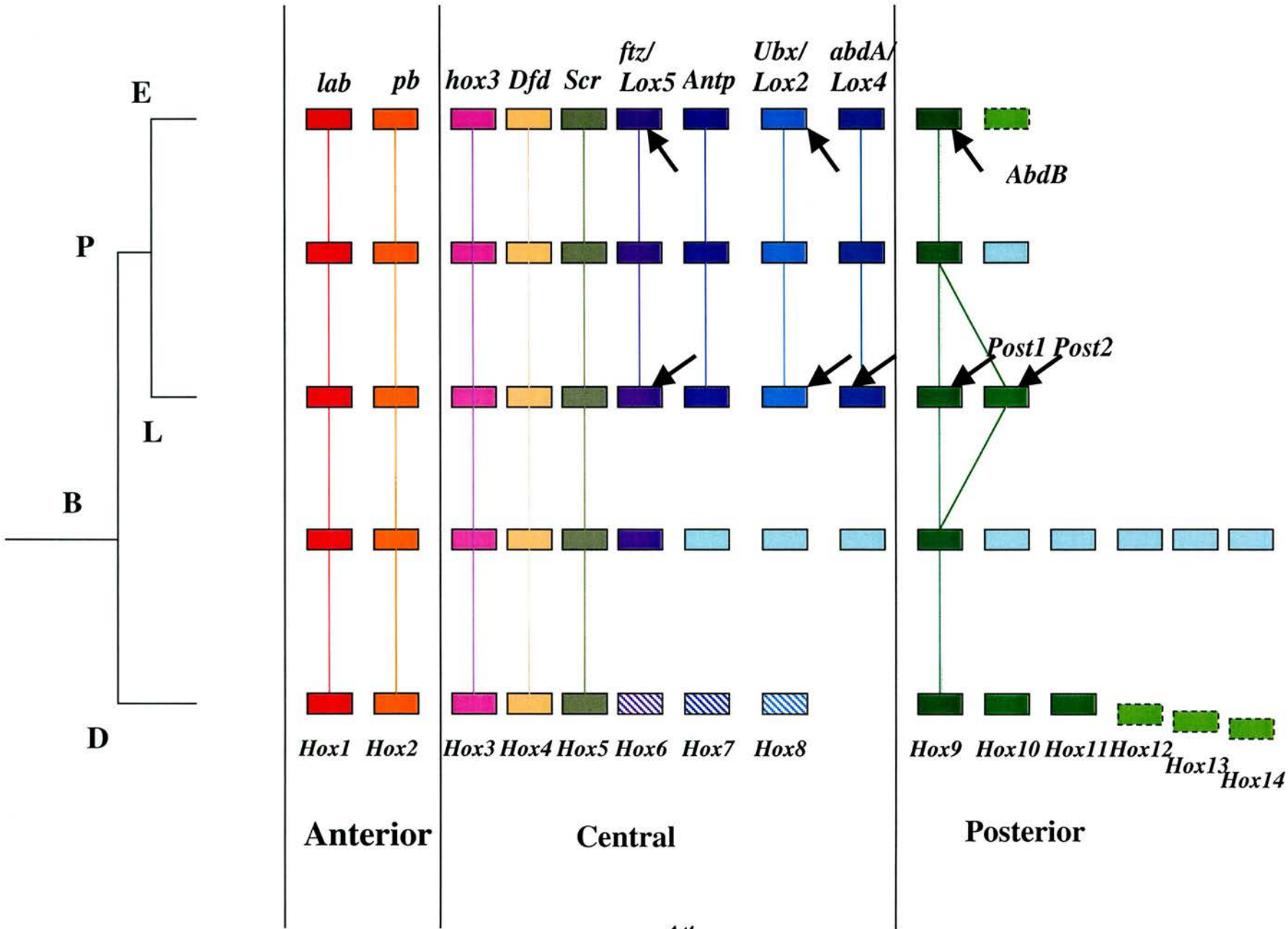


Fig 1.1 Tracing the evolution of the Hox cluster through the Bilateria

Each node in the tree represents a different point in evolution of the Bilateria. The extant basal Hox cluster of the Ecdysozoa (E), Lophotrochozoa (L), and the Deuterostomata (D) are estimated by a comparison of Hox gene complements form extant phyla (de Rosa et al., 1999). Orthology groups are colour coded and joined by lines if orthology is certain. Light blue boxes indicate genes that may have been present if the maximum number of Hox genes is assumed and then subsequently lost in some lineages. Boxes enclosed by dashed lines indicate genes present in some taxa of a clade that may or may not be ancestral. Black arrows pointing at protostome genes indicate groups that contain signature peptides and residues in and flanking the homeodomain, supported by phylogenetic analysis. Boxes patterned by stripes indicate genes of uncertain orthology in comparison to other groups.

The first five Hox cluster genes are orthologous across all three clades and were therefore present in the ancestral Bilaterian. Orthology between the central group genes of the Deuterostomata and the Protostomes has either been obscured by sequence evolution or alternatively these genes were not present in the basal Bilaterian. In this case they have arisen independently by independent duplications from one central gene. At least one posterior group gene was present in the ancestral Bilaterian but if the state of an expanded number of posterior genes observed in Deuterostomata is ancestral (the extreme case being Amphioxus with 6 posterior genes) it could be as many as six, with subsequent loss in the Protostomes, vertebrates and echinoderms. This places the minimum number of genes at 7 and the upper limit at 14.

The ancestral protostome may have had at least 10 Hox genes if the central gene groups represented by *Ubx/Lox2* and *abdA/Lox4* are orthologous and did not arise by independent duplications in each lineage. If the presence of two distinctive posterior genes in the Lophotrochozoa, and the presence of other divergent posterior genes in the two Ecdysozoan taxa (nematodes and priapulids) as well as the characteristic *AbdB* gene is ancestral, then there may have been as many as 11 genes in the basal protostome.

The existence of Hox gene signatures in the two protostome clades has been used as independent molecular evidence for the alliance of taxa to one clade or another (de Rosa et al., 1999). These have also been used to ally groups of previously uncertain affinity to one or other protostome group. For example, Kobayashi et al reported the cloning of a *Lox5* gene with the conserved *Lox5* peptide from a dicyemid. This group were previously considered as being basal and primitive because they have very simple body plans. The discovery of a Lox5 gene suggests that in fact the simple morphology of this group is derived, resulting in the loss of the more complex morphology of the Protostomes (Kobayashi et al., 1999). Hox gene complements and the signatures they contain will continue to be used as independent molecular data to support the inclusion of taxa within major clades.

However, these Hox signatures remain unpolarised in the absence of Hox genes from either an extant Bilatarian clade that is clearly basal to all the other Bilaterian groups or an outgroup to the Bilateria close enough to have an informative Hox complement. For example, the only Hox gene data supporting the inclusion of the Nematoda in the Ecdysozoa is the presence of an *AbdB* posterior gene (Fig 1.2). If the *AbdB* state is actually ancestral to the Bilateria than the presence of this gene would not unequivocally support the inclusion of a taxa in the Ecdysozoa. It is only in the light of other data that the Nematoda can be included in this group and the presence of an *AbdB*-like gene only supports this. The same argument also holds true for the discovery of the *Lox5* gene in the dicyemid (Kobayashi et al., 1999). Two candidates for the prestigious title of Bilaterarian outgroup are the Acoela, part of the platyhelmimthes now widely considered to be polyphyletic (Ruiz-Trillo et al., 1999), and possibly the Gnathostomulida as suggested by the combined molecular and morphological study described above (Peterson and Eernisse, 2001).

1.4. The special case of the *C. elegans* Hox cluster

The most dramatic exception to the conservative nature of Hox gene evolution is seen in the model organism *C. elegans*. Originally the small cluster of only four genes was believed to correspond to a primitive state of an unsegmented member of

the Bilateria (Burglin and Ruvkun, 1993; Burglin et al., 1991; Kenyon, 1994; Kenyon et al., 1997). However, the new picture of animal phylogeny places the Nematoda in with the Arthropoda in the Ecdysozoa (Aguinaldo et al., 1997; Blaxter, 1998). The completion of the *C. elegans* genome discovered two more Hox genes both belonging to the posterior class of Hox genes (Ruvkun and Hobert, 1998). One of these genes, called *php-3* (*posterior hox paralog 3*), is closely related to the posterior *AbdB* gene found in Arthropods and the other Ecdysozoa studied to date (de Rosa et al., 1999). Along with other molecular phylogenetic data this provides strong evidence that Nematoda are part of the Ecdysozoa.

However, closer inspection of the *C. elegans* Hox cluster suggests that the amount of modification since the last common ancestor of the Ecdysozoa is enormous. Firstly, by comparison with presumed ancestral Ecdysozoan Hox cluster shows that only four orthology groups are represented (*ceh-13=Hox1/lab*, *lin-39=Dfd*, *mab-5*=one central gene group probably *ftz/Lox5* or *Antp*, *php-3=AbdB*, *egl-5/nob-1*=posterior orthology group genes of unknown origin) and upto six may have been lost (*Hox2/pb*, *Hox3*, *Hox5/Scr*, *ftz/Lox5* or *Antp*, *Ubx/Lox2*, *abdA/Lox4*) (Finnerty and Martindale, 1998). Secondly, the Hox genes are not closely linked in a cluster of genes, but are instead spread through the genome with many non-Hox genes in between (Fig 1.3 A and B). There has also been an inversion in the cluster causing the central Hox gene *lin-39* to be located further 3' than the anterior Hox gene *ceh-13* (Fig 1.3). Finally, the remaining *C. elegans* Hox genes have divergent homeodomains that appeared to have evolved significantly from those of other protostome phyla ((de Rosa et al., 1999).

This arrangement suggests that *C. elegans* development no longer requires the majority of Hox genes and also that it no longer requires its Hox genes to be tightly clustered or to obey spatial colinearity. From this description it is clear why the *C. elegans* was originally thought to represent an ancestral state. Some radical changes must have occurred during the evolution of developmental mechanisms in *C. elegans* to allow the loss of the majority of Hox orthology groups. The development of *C. elegans* is characterised by an invariant cell lineage controlled by the unequal separation of cell fate determinants that first occurs by the unequal first cell division to form non-equivalent blastomeres (Sulston et al., 1983). In this

situation it is possible to see how the requirement for Hox genes and other developmental control genes involved in providing positional identity might be lost. The requirement for cells to be told where along the body axis they are may be lost if they are able to gather all the information they need by knowing which lineage they come from.

In order to investigate the evolution of developmental mechanisms that has led to the state observed in *C. elegans* it would be useful to know when the in the evolution of the *C. elegans* lineage the observed changes occurred. If the loss of Hox genes is related to the cell-lineage mode of development and if this mode is prevalent throughout the Nematoda then the observed Hox gene loss may have occurred for the whole phylum. The early development of some nematodes only distantly related to *C. elegans* has been investigated, and although the overall organisation of embryogenesis is very similar early cell-lineage is not predetermined (Cunha et al., 1999; Voronov and Panchin, 1998; Voronov et al., 1998). In the nematode *Enoplus brevis*, a member of clade II nematodes, the first cell division is symmetrical with either blastomere able to form posterior or anterior structures (Voronov and Panchin, 1998). So if the adoption of a determined cell lineage developmental mode is linked to Hox gene loss, nematodes only distantly related to *C. elegans* may have maintained Hox genes absent from *C. elegans*. Alternatively the Hox gene loss observed in *C. elegans* could be representative of the whole phylum and may have occurred before the radiation of the nematode lineages (see chapter 2). Whatever the exact nature of the Hox cluster in nematodes, comparative studies of the functions of Hox genes between *C. elegans* and the closely related freeliving nematode *P. pacificus* indicate that Hox gene function is evolving rapidly (see below).

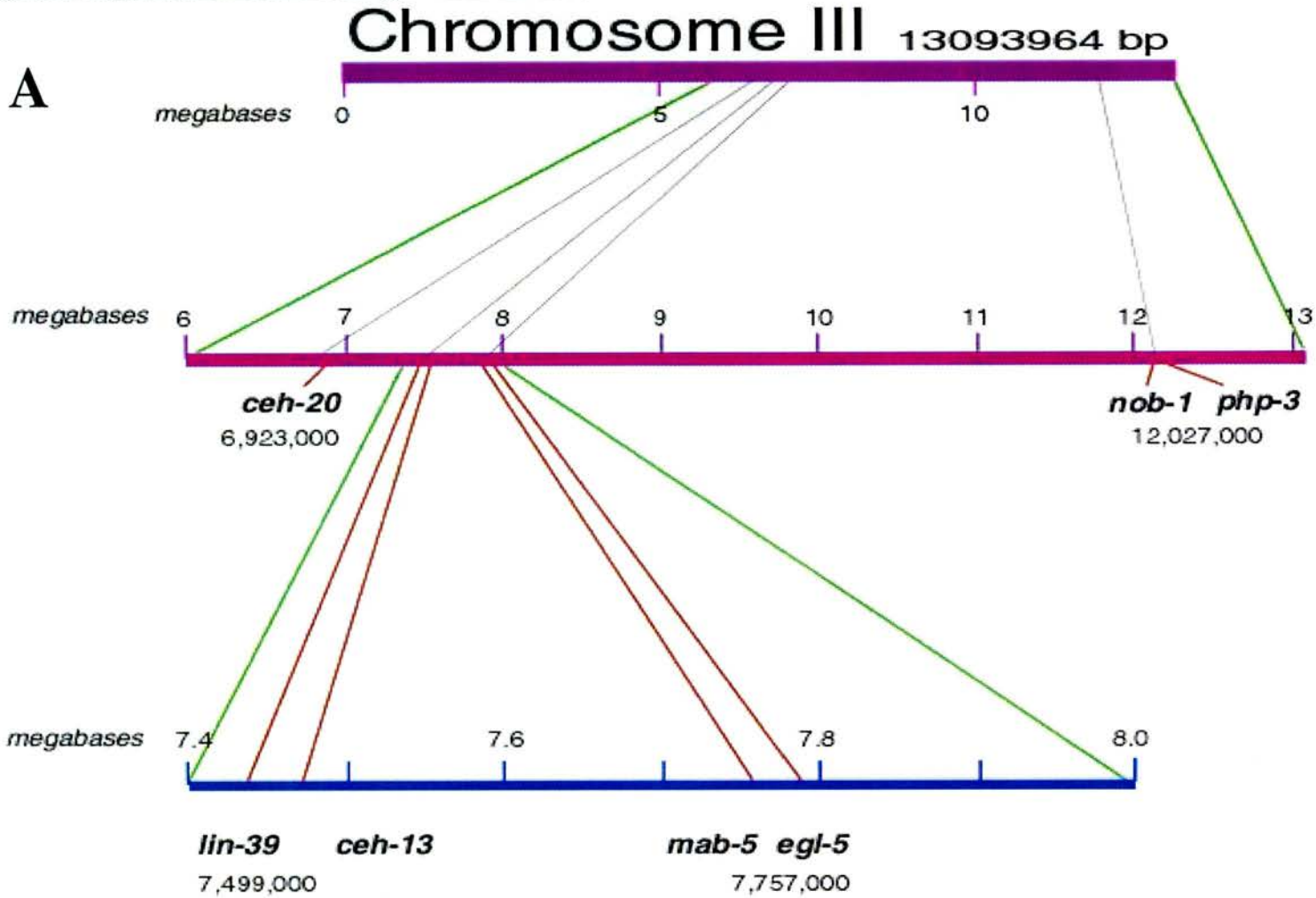
1.5. Outline of Hox gene functions in C. elegans

The function of all the *C. elegans* Hox genes has been investigated in some detail (Kenyon et al., 1997). These studies indicate that three of the Hox genes, the anterior group gene *ceh-13* and the posterior group genes *php-3* and *nob-1* are required during embryogenesis (Brunschwig et al., 1999; Van Auken et al., 2000; Wittmann

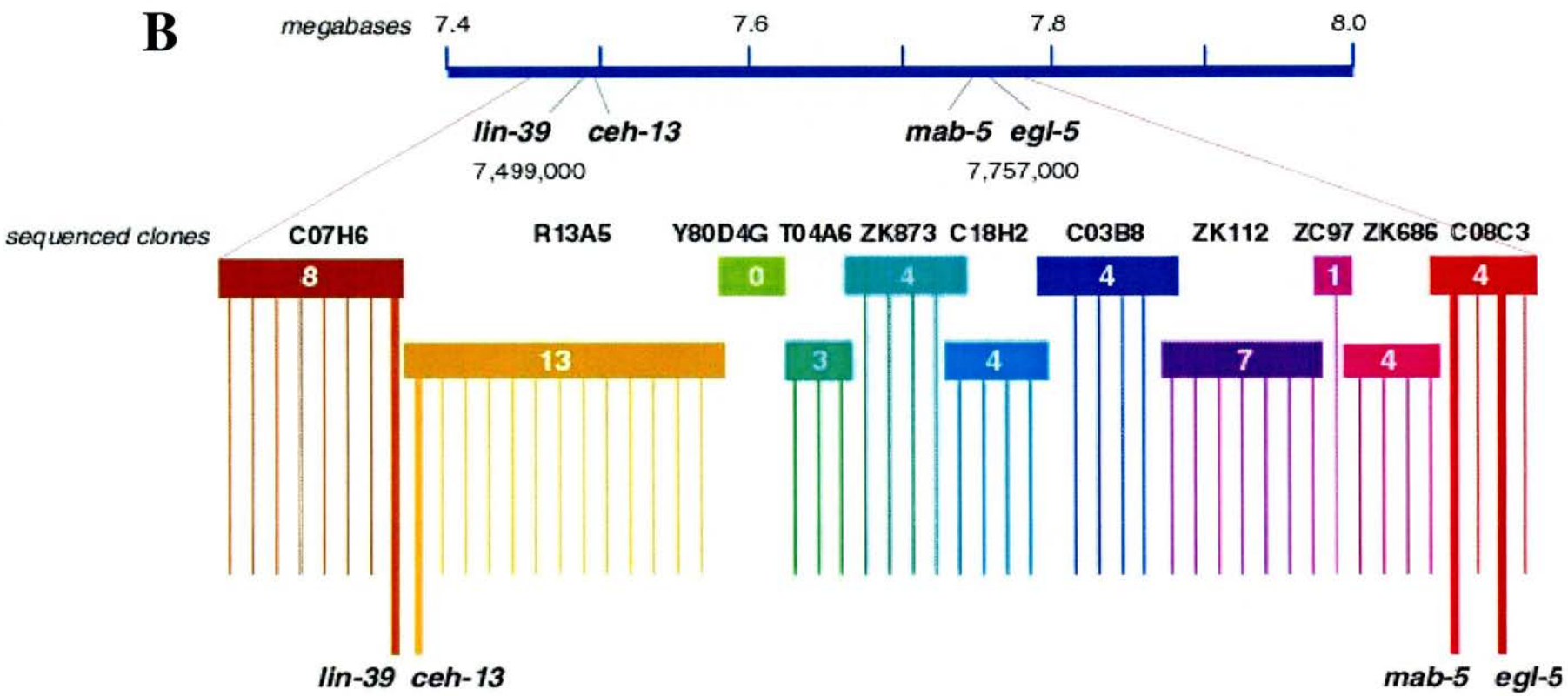
et al., 1997b). The remaining three genes are not required during embryogeneis, despite being expressed in some cells in the later stages of embryogenesis, and instead are required for the specification of cell fates and cell migrations in larval stages (Chisholm, 1991; Clark et al., 1993; Ferreira et al., 1999; Maloof and Kenyon, 1998; Salser and Kenyon, 1996).

The *ceh-13* gene is expressed in the anterior of the *C. elegans* embryo as would be expected by its orthology grouping, but not by its chromosomal position (Wittmann et al., 1997a). Its most posterior region of expression is located anteriorly to those of the other *C. elegans* Hox genes. Expression of the *ceh-13* gene is controlled by cell intrinsic mechanisms and is dependent on cell-linage and independent of cell position along the body axis (Brunschwig et al., 1999).

Fig 1.3. Organisation of the *C. elegans* Hox cluster



The *C. elegans* HOX cluster on Chromosome III



HOX and non-HOX genes in the centre of the *C. elegans* cluster

Fig 1.3 Organisation of the *C. elegans* Hox cluster

An overview of the genomic structure of the *C. elegans* Hox cluster spread over chromosome III. **A** The *C. elegans* Hox cluster is spread over approximately 5 Mb of chromosome III. The *lin-39* and *ceh-13* genes are inverted along the chromosome with respect to the law of spatial colinearity. This arrangement suggests a minimum of at least one chromosomal inversion and one movement of the two posterior genes *php-3* and *nob-1* away from the centre of the chromosome. Alternatively the nob-1 gene may have arisen as a recent duplication once the php-3 gene had moved. **B** An expanded view of the central part of the *C. elegans* Hox cluster spread over approximately 250 kb. Coloured boxes indicate sequenced cosmids spanning the region and the numbers in the boxes indicate the number of open reading frames identified by the *C. elegans* genome project. There are 39 predicted and confirmed genes between the two Hox gene pairs. In addition there is a gene between the *mab-5* and *egl-5* genes.

ceh-13 is therefore not receiving positional information like Hox genes in arthropods and vertebrates but is controlled by cell fate determinants and cell polarity cues. Analysis of the *ceh-13* mutant phenotype indicates that the protein is required for the proper organisation of the anterior of the embryo, and in particular its absence results in cell adhesion defects in this region (Brunschwig et al., 1999).

The other two genes required during embryogenesis pattern the posterior of the embryo. Elimination of the function of *php-3* and *nob-1* (*no-back end 1*) results in gross embryonic defects in the posterior of the embryo in both the pattern of cell divisions (something not directly effected in *ceh-13* mutants) and posterior cell migration events (Van Auken et al., 2000). In addition their absence causes posterior to anterior cell fate transformations as shown by the expression of anterior markers in more posterior regions than in wildtype worms (Van Auken et al., 2000). It is unknown whether these transformations are position specific or lineage specific as is observed for *ceh-13* and *mab-5* mutants (see below). Interestingly the *AbdB* ortholog *php-3* appears to have only a minor role in posterior patterning compared with the role of the more divergent *nob-1*. Use of RNAi to assess the effects of loss of function of these genes independently indicate that the *nob-1* null phenotype is much more severe than that of *php-3* (Van Auken et al., 2000). The double mutant is more severe still indicating that these genes act together to pattern the posterior of the embryo (Van Auken et al., 2000).

The third posterior gene in the *C. elegans* genome, *egl-5*, is involved in specifying cell fates in the posterior region of the body including the tail region of both sexes including the male reproductive organs (Chisholm, 1991; Ferreira et al., 1999). The phenotype of the *egl-5* mutation is post-embryonic suggesting that the gene has no essential role during embryogenesis (Chisholm, 1991; Ferreira et al., 1999). Once *egl-5* expression is established in a tissue it remains turned on throughout development. The role of *egl-5* seems to be to specify correct cell attachments and cell migrations, particularly in the presumptive posterior gonad of hermaphrodites and the male tail copulatory rays. Some evidence is also available to suggest the *mab-5* gene is required to initially establish *egl-5* expression in the mail tail ray lineages and posterior seam cells (Scott Emmons personal communication),

and the region upstream of *egl-5* in the genome contains sites that could be bound cooperatively by *mab-5* and *ceh-20* (a *C. elegans* PBX family gene).

The *mab-5* gene is required for proper development of posterior cell lineages into a variety of structures including lineally unrelated sensory organs (the male tail), muscles and epidermal structures located in the posterior body region (Salser and Kenyon, 1996). Like the *egl-5* gene it is also not required embryonically and expression of *mab-5* is lineage not position dependent (Cowing and Kenyon, 1996). Control of *mab-5* expression is complex because the cells expressing the gene do so in a lineage dependent manner rather than according to their position and they come from a number of different lineages (Harris et al., 1996; Kenyon et al., 1997). Other cells originating from the same lineages do not express *mab-5* (Kenyon et al., 1997). The *mab-5* expression domain overlaps with that of the *lin-39* gene (Kenyon, 1994). In this overlap they are known to be jointly involved in patterning two sets of cells, the M lineage that gives rise to the post-embryonic mesoderm (Liu and Fire, 2000) and the Pnp cells a subset of which are induced to form the vulva by the *lin-39* gene (Maloof and Kenyon, 1998). The combinatorial action in cells of the M lineage is dependent on interaction with the *ceh-20* PBX homolog (Liu and Fire, 2000).

The role of *lin-39* is perhaps the best characterised as vulval development in *C. elegans* is one the most explored genetic pathways (Maloof and Kenyon, 1998; Maloof et al., 1999). The role of lin-39 in vulval formation has two stages Firstly, *lin-39* is required early in the first larval stage to specify the vulval precursor cells (VPCs) and stop them undergoing cell fusion with the hypodermal sycitium. Secondly, in the third larval stage *lin-39* directs vulval cell fates switching on genes that start the process of vulval morphogenesis in response to inductive EGF/RAS/MAPK signalling (Maloof and Kenyon, 1998).

A consideration of Hox gene function in *C. elegans* suggests many similarities too some of the details of Hox gene function in vertebrates and arthropods. Their roles in specifying cell fates along the body axis obey spatial colinearity, even if their chromosomal order doesn't. In addition their requirement for the PBX homolog CEH-20 has also been demonstrated (Liu and Fire, 2000). However, three of the genes are not required embryonically and cells that express

Hox genes do so not because of their position along the body axis but because of the cell lineage (Cowing and Kenyon, 1996). Perhaps in the evolution of the nematode lineage, and in particular *C. elegans* and the other freeliving nematodes of small cell number, the remaining Hox genes have been coopted into a developmental mode largely dependent on asymmetric separation of cell lineage determinants and short range cell-cell interactions. This may obviate the need for Hox genes, or the presumed ancestral role of Hox genes to define position along the body axis.

A study of the Hox gene complements in nematodes only distantly related to *C. elegans* presumed to predominantly depend a cell lineage mode of development and others that may not, could provide evidence for this proposal. At this point this classification of whether nematodes distantly related to *C. elegans* have a predetermined cell lineage is based only on the observation of early cell divisions to see if asymmetry occurs early in embryogenesis (Voronov and Panchin, 1998) or on an investigation on the prevalence of eutely (Cunha et al., 1999). Without broader sampling of the Nematoda it will remain unclear how prevalent a determined cell lineage mode of development is in the Nematoda. Howver it is difficult to see how larger nematodes with larger cell numbers could avoid the use broader domains of positional information. So the larger nematodes present in some nematode clades (like the entirely parasitic clade III, see Fig 2.1) may have retained more Hox genes with the function of providing positional information in cell a lineage independent manner.

1.6. Comparative studies of *mab-5* and *lin-39* in P. pacifcus

Further evidence that Hox genes might be evolving rapidly comes from studies in the satellite model nematode *P. pacficus,* a freeliving nematode relatively closely related to *C. elegans* (Sommer, 1997a; Sommer, 1997b). Studies of *mab-5* and *lin-39* in this species (*Pp-mab-5* and *Pp-lin-39*) indicate while they are playing very similar roles in both species the are doing them in very different ways (Eizinger et al., 1999; Eizinger and Sommer, 1997b; Jungblut and Sommer, 1998; Sommer et al., 1998; Sommer and Sternberg, 1996).

Ce-LIN-39 has two distinct roles to during vulval formation as described above. Firstly it specifies the vulval equivalence group or VPCs then is required later to facilitate induction of the VPCs to form the vulva. During the first stage it specifies the VPCs by preventing these cells from undergoing fusion with the hypodermal syncitium, which is the default pathway of the surrounding cells. In the absence of *Ce*-LIN-39 the cells that normally become VPCs also undergo cell fusion (Clark et al., 1993). Use of heat shock contructs in a *lin-39* mutant background clearly shows that *Ce*-LIN-39 is then required to induce the vulval fate in response to a signal from the anchor cell of the gonad (Maloof and Kenyon, 1998). In *P. pacificus Pp*-LIN-39 also specifies the VPCs and the surrounding cells without *Pp*-LIN-39 undergo programmed cell death as oppose to cell fusion (Eizinger and Sommer, 1997b; Sommer et al., 1998). However, *Pp*-LIN-39 is not required for subsequently for vulval induction in *P. pacficus* as double mutants of *Pp-lin-39* and *Pp-ced-3* (a gene required for programmed cell death) develop. Thus it appears that in *P. pacifus Pp-lin-39* works in a different genetic pathway from *Ce-lin-39*, inhibiting programmed cell death. It has also been shown that the *Pp-lin-39* coding region will rescue the *Ce-lin-39* mutant, indicating that the changes that have allowed cooption of the *Pp-lin-39* gene have occurred in regulatory rather than coding regions (Grandien and Sommer, 2001). As cell fusion is observed in nematodes that are an outgroup to these to nematodes it is assumed that cell death is the derived state for this process (Eizinger and Sommer, 1997b; Felix et al., 2000).

Similarly the function of *Pp-mab-5* gene is similar to that seen in *C. elegans*. However, there are subtle differences in the role of these two genes in patterning the posterior PnP cells (the cells that give rise to the VPCs). *Pp-mab-5* is required to prevent the differentiation to a vulval cell fate of a posterior PnP cell that normally doesn't assume a vulval cell fate (Jungblut and Sommer, 1998). This is the same role played by *Ce-mab-5* except that in *C. elegans* this achieved by reducing competence to an inductive signal while *Pp-mab-5* acts to prevent a vulval cell fate that is independent of an inductive signal (Jungblut and Sommer, 1998). Although this doesn't necessarily indicate that evolutionary changes have occurred in the *mab-5* genes themselves it does at least mean that the two genes are functioning in different

genetic circuits as like *Pp-lin-39* the *Pp-mab-5* gene does not seem to be required to interpret the inductive EGF/RAS/MAPK signal (Eizinger et al., 1999).

These comparative studies indicate that significant functional evolution has occurred in these two Hox genes even between two relatively closely related species. It seems that the exact roles of Hox genes within the genetic networks of a deterministic mode of development are flexible. If the idea that Hox genes maintained in nematodes have evolved or evolving functions within a developmental framework radically from the one in which they were present ancestrally, perhaps we would expect them to assume roles in different genetic circuits in different species, even if the species are relatively closely related.

CHAPTER 2: HOX GENES FROM THE PHYLUM NEMATODA.

2.1. Choice of taxa for investigating Hox gene evolution in the phylum Nematoda

Accepting the current view of animal phylogeny, based on the synthesis of recent molecular phylogenetic and morphological data, means that the *C. elegans* Hox cluster has a derived structure when compared to other animal phyla (Adoutte et al., 2000; Aguinaldo et al., 1997; de Rosa et al., 1999; Peterson and Eernisse, 2001). It is missing orthologous groups present in the ancestral lineage, which importantly have therefore been lost. Molecular phylogenetic analysis of the remaining Hox genes indicates that they have undergone significant sequence evolution within the homeobox region compared to their invertebrate and vertebrate homologs while retaining signature peptides that allow their assignment to orthology groups or at least to a set of paralogs (Burglin and Ruvkun, 1993). An analysis of Hox gene function in *C. elegans* indicates some nematode Hox genes have only post-embryonic functions in contrast to some other animal phyla (Kenyon, 1994), which may be a result of redundancy created by the cell lineage mechanism of *C. elegans*. The weight of evidence also suggests that Hox gene expression in *C. elegans* is essentially cell lineage dependent and not directly related to position along the body axis (Brunschwig et al., 1999; Cowing and Kenyon, 1996). This supports the hypothesis that Hox genes have been coopted into roles within a predominantly deterministic mode of development. Finally, comparative evolutionary development analysis of Hox genes and the genetic networks within which they function in the "satellite model" *Pristionchus pacificus* suggests that Hox genes are undergoing rapid evolution at least in this part of the nematode lineage (Eizinger et al., 1999). These points summarise those discussed in detail in the last section of the previous chapter.

The importance of Hox genes in the evolution of animal body plans is unquestioned but its exact nature still remains unclear (Finnerty and Martindale, 1998). In order to better understand the molecular basis of Hox gene evolution a larger number of data points encompassing close, medium and distant phylogenetic

relationships need to be studied in the context of different evolutionary histories. The nematodes are relatively un-sampled phyla with the only available Hox gene data from two relatively closely related free-living nematodes. There are a number of reasons for this despite the ubiquity and socio-economic importance of nematodes, and the existence of *C. elegans* model as an ideal starting point for comparative analysis (Aboobaker and Blaxter, 2000; Blaxter, 1998). Perhaps the basic body plan of nematodes considered as simple, the existence of entirely parasitic clades (Blaxter et al., 1998), the small size of free-living species and the attractiveness of the Arthropods based on the same criteria are important reasons why nematode Hox genes have not been sampled. Sampling a cross section of the whole phylum would help to answer the first set of questions associated with evolution of the *C. elegans* Hox cluster and the function of these genes. When did the observed gene loss occur? Was it before or during the origin of the nematode lineage? When did the observed sequence evolution within the divergent homeodomains occur? In an attempt to answer these questions a number of species have been chosen as representatives of a cross-section of the phylum.

A strong molecular framework for the phylogeny of nematodes is now available (Fig 2.1) (Blaxter et al., 1998). This has been used to select the six most appropriate nematode species for investigation, representing a spectrum of distant to very close relationships to *C. elegans*. This will allow the investigation of the evolution of the nematode Hox cluster over different evolutionary time spans. They have also been selected for their relationships to each other, independent of *C. elegans,* so that the mode and tempo Hox gene evolution of over the whole phylum can be measured (i.e. not just in comparison to *C. elegans*). Other advantages in the choice of these species include the existence of relevant data and the availability of many essential resources, most importantly fresh nematode samples. These six species are *Pristionchus pacificus, Brugia malayi, Ascaris suum, Strongyloides ratti, Meloidegyne javanica* and *Trichinella spiralis.* These taxa represent members of 4 of the 5 major clades of the phylum, clade II has not been sampled yet (Fig 2.1). Comparison between the Hox gene complements of these taxa will allow the direction, tempo and mode of evolution of the Hox cluster in nematodes to be assessed.

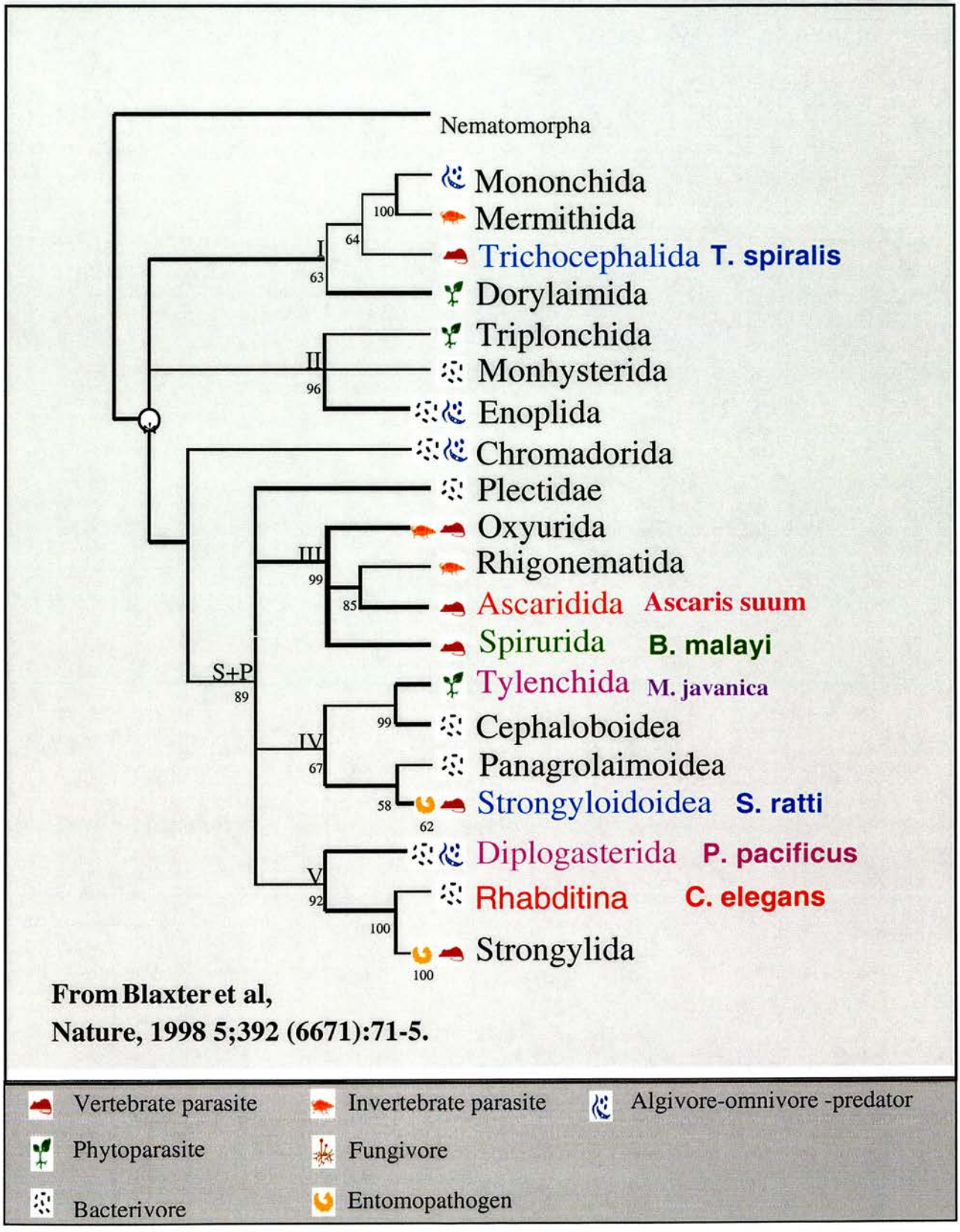
The first nematode chosen for study was the filarial parasite *Brugia malayi,* which is the subject of an advanced genome project co-ordinated by the World Health Organisation (Williams et al., 2000). A more detailed description of this representative of a group of major human parasites is provided in chapter 5 of this thesis. The genome project has provided a large number of resources for the filarial research community including over 23,000 ESTs, a selection of cDNA libraries prepared from points around the whole lifecycle and gridded bacterial artificial chromosome (BACs) genomic libraries (Williams et al., 2000). Thus this nematode species represents a good candidate for studying conserved gene families, such as the Hox genes, outside of the model nematode *C. elegans*. It was anticipated that the resources available would not only aid in the initial discovery of Hox genes from this species but would also allow further characterisation to proceed relatively rapidly.

Brugia malayi is only distantly related to *C. elegans* and is a member of a large clade that consists of entirely parasitic species of both vertebrates and invertebrates (Fig 2.1). Data about its Hox gene cohort therefore offers a comparison of between free living and parasitic members of the same phylum. Hox genes from parasite species have not been studied extensively in animal phylogeny. Although *a priori* adaptive evolution of the anteroposterior patterning of the body plan might not be expected to be affected directly by a parasitic mode of life compared to any other there are suggestions that parasitic species tend to lose genes in comparison to there free living relatives. This is perhaps based on the assumption that implicit in the parasitic lifecycle is the use of host mechanisms and resources. It can be argued that this leads to the loss of structures and functions adapted for free-living lifecycles . In the case of Hox genes there is one possible example of Hox gene loss that may be connected with body plan adaptations to the parasitic mode (Mouchel-Vielh et al., 1998).

Brugia malayi and the parasitic clade it represents have, for the most part, body plans consistent with the relatively simple nematode body plan. They tend to be much larger than the free-living nematodes like *C. elegans* and its close relatives. There are also some notable differences that are probably adaptations to a parasitic mode, like the flexible proteinaceous larval sheath that encapsulates the first stage larvae (Smith, 2000). Some differences like the position of the vulva (placed to the

anterior in *Brugia malayi*) might be, by analogy with *C. elegans,* due to differences in Hox gene expression. As previously discussed one possible reason for the evolution of a reduced and separated *C. elegans* Hox cluster is the employment of a highly cell-lineage dependent mode of development. Such a mode of development may make the role of Hox genes in patterning the anteroposterior axis partially redundant, as it may now be no longer dependent on the presence or absence of the activity of genes that control the formation of region specific structures along the body axis. The success of this almost invariant cell lineage dependent mode of development in free-living nematodes could be attributed in part to their relatively small size and small cell number. Larger animals with higher cell number paired with a large number of specialist tissues and structures might not be able to depend entirely on cell lineage mechanisms, requiring instead systems that indicate relative spatial positions of cell groups (tissues). This may be true for *Brugia malayi*, and the parasitic clade it represents, as it is a much larger organism than *C. elegans.* However, alternatively it is possible that a cell lineage mode of development is still used during embryogenesis of these larger nematodes and increased size is just a matter of having more of the same cells. If this were true than it could be the relative simplicity of the nematode body plan and not size that is important for the use of a cell lineage mode of development. Analysis of the Hox genes in this nematode and other species would be a good starting point in order to address this interesting question. The points discussed here are thus relevant for all the species included in this study.

Fig 2.1. An overview of the phylum Nematoda



From Blaxter et al,
Nature, 1998 5;392 (6671):71-5.

Fig 2.1. An overview of the phylum Nematoda

A robust molecular evolutionary framework based on the 18rDNA gene exists for the phylum (Blaxter et al., 1998). The Nematoda are separated into five major clades numbered I-V. The tree highlights the names of the major groups in each clade and indicates their trophic ecology. Clade V includes the model organism *C. elegans* and free-living satellite model *P. pacificus*. The other nematodes chosen from which Hox genes have been identified in this study are *M. javanica* (IV), *S. ratti* (IV), *A. suum* (III), *B. malayi* (III) and *T. spiralis* (V).

2.2. Cloning of Hox genes from a cross section of the phylum Nematoda by using a two step degenerate PCR and subtractive screening approach.

In order to allow the rapid identification of Hox genes from nematodes a degenerate PCR approach widely used in other studies was employed (Balavoine and Telford, 1995; de Rosa et al., 1999; Dick and Buss, 1994b; Pendleton et al., 1993; Ruddle et al., 1994b). Although the primer sets used would be expected to amplify all *Antennapaedia* class homeodomains many previous studies have shown this not to be the case (Cook et al., 2001; de Rosa et al., 1999). The reasons for this are not clear but are likely to be the result of a number of factors. One possible explanation is that some genes may have divergent homeodomains in the regions normally amplified. Analysis of the available nematode Hox gene sequences suggested this could be a particular problem for this phylum. In order to deal with this problem novel degenerate primer sets were employed (see below). Another problem for a degenerate PCR approach when using cDNA as a template is that some genes will have higher expression levels than others and are therefore likely to represent the majority of subsequently cloned PCR products. In order to combat this a subtractive PCR screen was used after cloning and initial sequencing to identify rare inserts (see below). This approach for partial cloning of nematode Hox genes is discussed below while the precise experimental details are in the Materials and Methods section.

2.2.1. Degenerate PCR primer design

Many previous studies have designed effective primers for the amplification and cloning of all *Antennapaedia* family and related homeobox genes. Most of these primer sets are designed to amplify the region encompassed by the highly conserved LELEKEF and KIWFQN peptides (Table 1 in Materials and Methods for full description of all degenerate primers and corresponding peptide sequence). These primers were used to amplify short homeodomain fragments as described below and in detail the Materials and Methods chapter. However, the observation that the available Hox gene sequences from *C. elegans* and the two published sequences from *P. pacificus* were divergent suggested that even the full complement of primers

used from other studies would not amplify all nematode Hox genes. In an attempt to deal with this possible problem additional degenerate primer sets were designed to each nematode Hox gene orthology group present in *C. elegans*. Primers to each *C. elegans* Hox gene were designed to regions of homology with genes from other phyla in their proposed orthology group. In addition the *Ce-mab-5* and *Ce-lin-39* genes were aligned to their orthologs in *P. pacificus*, the sequences of which were already available (Jungblut and Sommer, 1998; Sommer et al., 1998). The *Ce-egl-5* gene could not be easily aligned to any Hox orthology group but did have a region of homology to other homeodomain proteins in the amino terminal of its homeodomain. In fact most of the identified regions lay towards the N-terminal of the homeodomain and provided alternative forward primers for amplifying these genes by degenerate PCR (See Table 1 in Materials and Methods for the sequences and names of these primers). Data from sequences generated from using these primer sets and subsequent extension of these clones by 5' or 3' RACE (**R**apid **A**mplification of **c**DNA **E**nds) were used iteratively to generate additional optimised primer sets for amplifying nematode Hox genes (Table 5.1 in Materials and Methods).

2.2.2.Cloning and sequencing of partial nematode homeodomains

Degenerate PCR with the primers designed above was performed with a range of PCR conditions on quantified nematode genomic DNA and cDNA from all the species chosen for study. The range of PCR conditions (see Table 2 in Materials and Methods) were chosen by using primers on *C. elegans* gDNA and cDNA, such that all Hox genes and some other *Antennapaedia* class genes (the *caudal* homolog *pal-1* and the engrailed homolog *ceh-16*) were amplified from first strand cDNA. In most cases quantities were limited so that gDNA and total RNA was prepared from the same material, with subsequent synthesis of first strand cDNA as described in the Materials and Methods section. In general amplification from cDNA proved to be highly successful, producing products of the predicted sizes. Amplification from gDNA was not as efficient as amplification from first strand cDNA. This was may have been due in some instances to low template quality and also the presence of large introns in the homeodomain of some nematode Hox genes. PCR products were

visualised by gel electrophoresis and gel purified before cloning. Inserts were sequenced and identified as partial homeodomains by using the BLAST family of algorithms to search public databases ((Altschul et al., 1997). Nucleic acid sequences were than translated into predicted protein sequences using the MacVector 7.0 (Oxford Molecular, UK) software suite and aligned using the ClustalW algorithm within this package.

2.2.3. A subtractive PCR screening approach to identify rare cloned inserts

Each degenerate PCR product was assumed to represent a mixed population of amplified products. This did not always prove to be the case, particularly for products that were generated by primers designed specifically to amplify nematode orthologous groups (described above). At least 96 and up to 960 clones were screened for inserts from cloning reactions of PCR products generated from cDNA or gDNA using primers from the "universal" group (see Table 1 in Materials and Methods). Of these between 24 and 96 clones positive for an insert of the correct size were chosen for preliminary sequencing. These sequences identified between 1 and 4 different partial homeodomain sequences. Insert specific primers were then designed and used in conjunction with vector primers to PCR screen all the remaining clones positive for inserts of the correct size. Two PCRs were performed on each clone with vector primers in opposite directions beginning with the most common insert identified by the preliminary sequencing. Any inserts positive by PCR were excluded from the next round of screening looking for the next most common insert identified by preliminary screening. Any remaining positive inserts, ranging from 1% to 15% of the total number of inserts, were than sequenced to identify the "rare" inserts not identified by preliminary sequencing. The number of "rare" inserts identified from a mixed population of products generated by degenerate PCR using this subtractive PCR approach ranged from 0 (all clones screened out) and 4.

For larger PCR products amplified from gDNA this approach was not used as it was assumed that these were likely to represent homeodomains containing an intron and therefore not a mixed population.

2.2.4. Extension of Hox gene sequences beyond the regions cloned by degenerate PCR

For some of the cloned Hox genes it was possible to obtain more sequence by using 5' and 3' RACE. An alternative method for extending Hox gene clones was to use inverse PCR from gDNA preparations when a suitable amount of material was available. This method was only attempted for *Brugia malayi* Hox genes. The details of these approaches are described in the next chapter and the in the Materials and Methods section. In some cases this extra sequence proved very informative in allowing the clearer assignment of genes to orthology groups. In this chapter only the sequences covering the extent of the homeodomain are presented. The sequences outside of the homeodomain, where known, are presented and/or discussed in the next chapter along with any refinements to analyses and discussions described here. All nucleic acid sequences generated during this work including some that are not discussed are presented in Appendix 1.

2.2.5. Assignement of nematode Hox genes to orthology groups

Here the nematode Hox genes are assigned to orthology groups using the presence and absence of characteristic residues and by their identity to other nematode genes identified. For the purposes of this analysis characteristic residues have been defined as those that are present in more than 50% of genes in an orthology group that are different from the *Drosophila melanogaster Antennapaedia* (*Antp*) gene homeodomain (after de Rosa et al, 1999). In addition to this nematode genes are grouped by identifying characteristic residues within the phylum that are shared between them, suggesting that they were orthologs. For this purpose nematode characteristic residues were those that were both different from *Antp* and from characteristic residues across phyla and present in the majority of nematode orthologs. This was especially important for some of the more derived homeodomain

sequences that could not be confidently placed into orthology groups by direct comparison to genes from other phyla.

2.3. An overview of Hox and Hox related genes from the phylum Nematoda

The degenerate PCR screen coupled with subsequent subtractive screening has identified a total of 41 partial homeodomains. Of these 41 genes 35 appear to be Hox genes based on initial assignment to orthology groups by using BLAST algorithms to search public databases (Altschul et al., 1997). Of the remaining 6 genes 5 are *caudal* family genes, the posterior group gene of the ParaHox cluster (Brooke et al., 1998) and one could not be assigned to any homeodomain group. Of the 35 Hox genes partially cloned 26 are definitely homologs of *C. elegans* Hox genes, another 9 genes are absent from *C. elegans* and are orthologous to Hox genes in other phyla. Among these genes at least three Hox orthology groups absent from *C. elegans* are represented, a *hox-3* group, at least one central group closest to the *hox 6-8* orthology groups of deuterosomes and to *fushi-tarazu/Lox5* or *Antennapaedia* in other protostomes, and a *Lox2/Ubx* central class gene. Described below briefly are the Hox genes found in each nematode species followed by a more detailed discussion of nematode Hox genes organised and aligned by orthologous groups in the next section. A phylogenetic analysis of nematode Hox genes is described in Chapter 4. Here alignments of orthologous groups are discussed with respect to the presence and absence of characteristic and/or unique residues within the sequences identified. All Hox gene nucleotide sequences are presented in Appendix 1.

2.3.1. An overview of Hox and Hox related homeodomains from Brugia malayi

Of all the species studied *Brugia malayi* had the most resources available for this study. A total of five cDNA libraries were screened from around different lifecycle stages of the parasite (Blaxter and Ivens, 1999). In addition to gDNA first strand cDNA was prepared from both gravid adult females and microfilaria (Mf, equivalent to L1 larvae of *C. elegans*) in non-limiting amounts for use in degenerate PCR. A

total of nine homeodomain genes have been cloned by the degenerate PCR approach. Extensive searching of the large EST dataset did not reveal any *Antennapaedia* class homeodomain proteins. It is not clear if this is because they are absent or if the portions that are present are in regions without homology to known genes. This is not unlikely, as all the ESTs generated from *Brugia malayi* have been generated using 5' reads from directionally cloned cDNA libraries (Blaxter and Ivens, 1999). The sequences may not reach the conserved homeodomain region outside of which very little conservation is observed for Hox genes (Gehring et al., 1994), except for the hexapeptide motif required to allow co-operative binding of DNA with the PBX family of genes (Passner et al., 1999; Piper et al., 1999a).

Of the nine genes partially cloned, seven appear to be Hox genes and two are homologs of the posterior ParaHox cluster caudal gene family. Of the seven Hox genes five are orthologs of *C. elegans lin-39, mab-5, ceh-13, egl-5 and php-3*. However, two of the *B. malayi* Hox genes are not present in the fully sequenced *C. elegans* genome. Initial assignment to Hox orthology groups indicates that one of these genes is a divergent central Hox cluster gene, an *antennapaedia* or *fushi-tarazu* ortholog. The second gene is even more divergent from Hox genes in the public databases. Assignment of this gene to an orthology group is thus difficult, but it appears to be marginally most closely related to the *Hox3* group. One of the two caudal homologs is the ortholog of the *C. elegans* gene *pal-1*. The second gene is not a *pal-1* homolog based on the small amount of sequence available. The three extra genes found in *Brugia malayi* therefore not only reflect additional genes compared to the model *C. elegans* but also two additional orthology groups.

2.3.2. An overview of Hox and Hox related homeodomains from Ascaris suum

Ascaris suum was made available by the ongoing parasitic EST project in Edinburgh. First strand cDNA and gDNA was prepared from adult female gonads dissected from the rest of the worm. *A. suum* is a vertebrate parasite from clade III of the phylum, and of the nematodes investigated here is most closely related to *B. malayi* (Fig 2.1). The degenerate PCR screen identified eight partial homeodomains. These

represented the seven hox genes also identified in *B. malayi* and also a caudal family gene. The caudal gene identified is an ortholog of the *C. elegans pal-1* gene rather than the second caudal gene identified in *B. malayi*.

2.3.3. An overview of Hox and Hox related homeodomains from Pristionchus pacificus

Pristionchus pacificus is a small free-living clade V soil nematode that has been utilised as a satellite model to *C. elegans* (Delattre and Felix, 2001). *P. pacificus* first strand cDNA and genomic DNA was prepared in bulk from a culture maintained in the laboratory, kindly provided by Professor Ralf Sommer from the University of Tuebingen. Professor Sommer's laboratory have worked extensively on Hox and Hox related genes from this species and the cloning of Hox genes from *P. pacificus* was performed in collaboration with his laboratory in the hope we might find Hox genes absent from *C. elegans*.

P. pacificus is placed in the same major clade as *C. elegans* and both the *Pp-lin-39* and *Pp-mab-5* have been studied extensively in the context of the evolution of cell fate determination between these two nematodes. The degenerate PCR screen identified a further four genes. These are homologs of the *C. elegans* genes *Ce-ceh-13, Ce-egl-5, Ce-php-3* and *Ce-pal-1*. The screen did not identify any genes absent from *C. elegans*, although it did identify homologs of all the *C. elegans* Hox genes (including *Pp-mab-5* and *Pp-lin-39*) with the exception of the divergent posterior group gene *Ce-nob-1,* which has not been found in any of the other nematodes.

2.3.4. An overview of Hox and Hox related homeodomains from Strongyloides ratti

Strongyloides ratti is a vertebrate parasite and is a member of clade IV (Fig 2.1). First strand cDNA and gDNA was prepared from adult females and infective larvae kindly provided by Dr Mark Viney at Bristol University and also by Professor Andrew Read and his group in Edinburgh. Clade IV of the Nematoda includes

animal parasitic, plant parasitic and free-living nematodes. A total of six Hox genes were partially cloned from this species representing orthologs of all the *C. elegans* Hox genes (except *nob-1)* and one extra Hox gene that appears to be a central group gene based on the short PCR fragment that was cloned. This gene is not obviously a homolog of the extra central gene identified in both *B. malayi* and *A. suum* and may represent an additional central orthology group in nematodes. An initial tentative identification is that it is a member of the *Ubx/Lox-2* group.

2.3.5. An overview of Hox and Hox related homeodomains from Meloidogyne javanica

Meloidogyne javanica is a plant parasite in the same major clade (IV) as *S. ratti.* First strand cDNA and gDNA was prepared from developing embryos and early larval stages from material provided by Professsor David Bird from North Carolina. A less extensive degenerate PCR screen of this nematode has found five Hox genes, including two copies of the *lin-39* gene. The 2 copies differ at one amino acid position in the overlapping region of their sequences and at 5 nucleic acid positions. The other three genes were homologs of *Ce-ceh-13*, *Ce-mab-5* and the putative *Hox3* homolog identified in both *B. malayi* and *A. suum.*

2.3.6. An overview of Hox and Hox related homeodomains from Trichinella spiralis

Trichinella spiralis is a parasite of vertebrate skeletal muscle and of the species sampled is the most distantly related to *C. elegans.* It is a member of clade I and as such is phylogenetically positioned close to the root of the Nematoda. Use of its more slowly evolving 18srDNA gene in phylogenetic analysis supports the Ecdysozoa hypothesis (Aguinaldo et al., 1997). In addition the full sequence of its mitochondrial genome is much more informative with respect to sequence and gene order as to the position of the Nematoda within the higher protostomes than the highly derived mitochondrial genomes of other nematodes (Lavrov and Brown, 2001). For this reason it was chosen as a good candidate for a comparative analysis

of Hox gene cohort and sequences with *C. elegans* and the other nematodes studied here. Both first strand cDNA and gDNA was prepared from gravid adult females available in small amounts and infective larvae. Material for these preparations was kindly provided by Professor Murray Selkirk from Imperial College, London and Professor Malcolm Kennedy from the University of Glasgow.

The degenerate PCR screen identified a total of nine partial homeodomain genes. These included orthologs of *ceh-13, egl-5*, *php-3*, the caudal gene *pal-1* and a probable *lin-39* ortholog. Two central class genes were also identified which are possibly homologs of *mab-5* and the extra central gene also found in *B. malayi* and *A. suum*. The possible *Hox3* homolog was also identified in *T. spiralis* and is much less divergent compared to non-nematode orthologs than those found in *B. malayi*, *A. suum* and *M. javanica*. This is important as it confirms that these genes do belong to the *Hox3* group despite being very divergent. Finally another divergent homeodomain that could not be placed in any orthologous group was also identified.

2.4. Orthology groupings of Hox genes from nematodes

Only four Hox gene orthology groups are represented in the Hox cluster of the model organism *C. elegans*. These are the *ceh-13/labial/Hox1* homolog , *lin-39* a *Dfd/Hox4*, *mab-5* most like the *Hox6-8/ftz/Lox5* and *antennapadia* groups present in other bilaterians, and the three posterior group genes *egl-5*, *php-3* and *nob-1* (Kenyon et al., 1997; Ruvkun and Hobert, 1998; Van Auken et al., 2000). Of these posterior genes *php-3* is certainly a homolog of the *Abd-B* posterior gene found in all members of the Ecdysozoa studied so far (de Rosa et al., 1999) while the other two divergent posterior genes are likely to be specific to nematodes (*egl-5*, see below) or just to *C. elegans* (*nob-1*). As previously discussed this arrangement was originally interpreted in the context of the hypothesis that *C. elegans* held a basal phylogenetic position with respect to other Bilaterian phyla and that its Hox cluster represented a primitive ancestral state (Burglin and Ruvkun, 1993; Burglin et al., 1991; Kenyon, 1994; Kenyon et al., 1997). However, with respect to the new picture of Bilaterian phylogeny and the delayed discovery of the *php-3* gene it must represent a derived structure (de Rosa et al., 1999). By comparison to other members of the Ecdysozoa

at least six Hox orthology groups must have been lost during the evolution of the lineage leading to *C. elegans*. These genes belong to the *Hox2/pb*, *Hox3*, *Scr/Hox5*, either *hox6-8/ftx/Lox5* or *antennapaedia* depending which group is represented by *mab-5*, *Ubx/Lox2* and *abdA/Lox4* orthology groups.

The survey of Hox genes from a cross-section of the phylum Nematoda has illuminated that at least three of these Hox genes have been lost from the *C. elegans* lineage after the expansion of the phylum into its 5 major extant clades (see Fig 2.1 and Fig 2.9). The four orthology groups present in *C. elegans* represented by six genes (*nob-1* has not been found in other nematodes) have also been found throughout the rest of the phylum, and in line with current convention for naming nematode genes have been named after their *C. elegans* orthologs. The three extra orthology groups are *Hox3*, represented by genes found in four species, and at least two (and possibly three) central gene orthology groups (one or two most like the *hox6-8/ftz/Lox5* or *antennapaedia* orthology groups and one most like the *Ubx/Lox2* group). The orthology groups present in the Nematoda are discussed below. The exact relationships of the new central genes to each other and to those in other phyla remains unclear on the basis of the alignments presented here and so these are considered together. More definite statements of their relationships to genes in other phyla are made in chapter 4 of this thesis, where molecular phylogenetic methods have been used to analyse nematode Hox gene evolution in comparison to other phyla.

2.4.1. Labial/Hox1/ceh-13 orthology group genes in nematodes

A *labial/Hox1* gene was identified from all the six species studied, although for most of them the known sequence only spans the centre of the homeodomain (Fig 2.2). An alignment of these genes with each other and genes of this group from other phyla confirms that all the nematode genes have some of the characteristic residues of this group. Nonetheless, these genes, like the *C. elegans* labial/Hox1 ortholog *ceh-13*, gene remain divergent from this class of genes in other phyla. For each nematode gene at least 2 characteristic residues have not been maintained across the region of

the homeodomain identified. By the criteria described above to identify nematode characteristic residues the nematode *Hox1* genes have 4 characteristic residues based on the available sequences. In addition there have been a large number of changes within the homeodomain of this gene during the evolution of the nematode lineage. Some of these changes are shared by more than one species and reflect the evolutionary relationship of the species based on the robust 18srDNA gene analysis available for this phylum (Fig 2.1). However, some shared residues can only be parsimoniously explained as convergent changes (Fig 2.2) to the same amino acid assuming that the robust 18srDNA phylogeny is correct. It appears that the rate of evolution of this orthology group within the Nematoda might outstrip that achieved by the whole of the rest of the Bilateria, although this possibility is not addressed formally until chapter 4 of this thesis.

Fig 2.2. Alignment of *Hox1/labial/ceh-13* genes

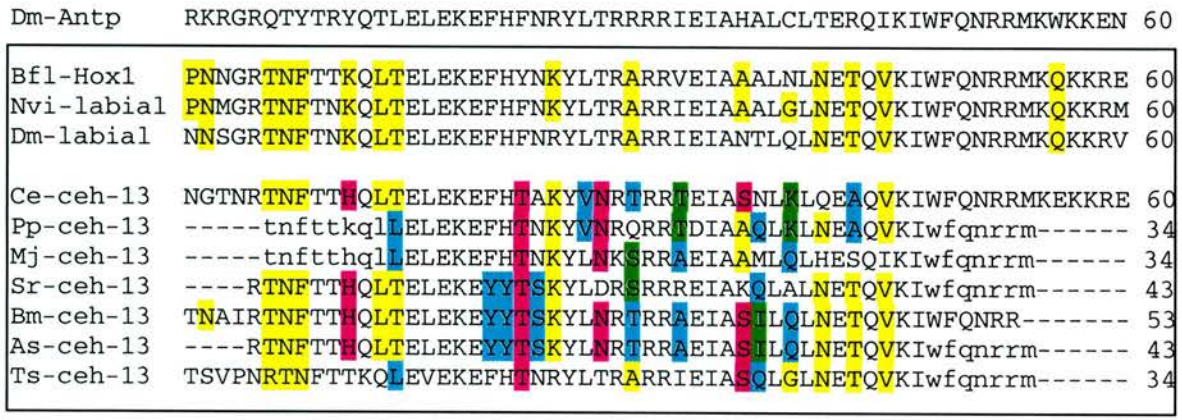


Fig 2.2. Alignment of *Hox1/labial/ceh-13* genes

An alignment of the nematode *ceh-13* genes with *labial/Hox1i* genes from other phyla compared with the *Drosophila melanogaster Antennapaedia* homeodomain. Yellow residues indicate characteristic residues for the orthology group defined as being different from the *Antp* homeodomain and present in more than 50% of known orthologs (after de Rosa et al 1999). Purple residues represent those likely to be characteristic of the nematode group by the same criteria for characteristic residues across phyla, different from both the *Antp* homeodomain and any characteristic residues for the orthology group. Turquoise residues are those different from the *Antp* homeodomain and any characteristic residues for the orthology group but shared by more than one nematode species. Numbers at the end of each line indicate the number of known residues in the homeodomain. Translations of the primer sequences used to amplify fragments are written in lower case when a fragment has not been extended by an alternative method.

Species abbreviations are Deuterostomata: Bfl, *Branchiostoma floridae*. Ecdysozoa: Dm, *Drosophila melanogaster*. Lophotrocozoa: Nvi, *Nereis virens*. Nematoda: Ce, *Caenorhabditis elegans*; Pp, *Pristionchus pacificus*; Sr, *Strongyloides ratti*; Mj, *Meloidogyne javanica*; Bm, *Brugia malayi*; As, *Ascaris suum*; Ts, *Trichinella spiralis*.

2.4.2. Hox-3 orthology group genes in nematodes

The *Hox3* orthology group is absent from the complete *C. elegans* genome and has therefore been lost during the evolution of this lineage. Four genes cloned in this study are likely to be members of this class (Fig 2.3). The sequences of these homeodomains for three of these genes, from the species *M. javanica*, *B. malayi* and *A. suum*, are very divergent from the *Hox3* group. Assessed individually, assignment of these sequences to this orthology group would be uncertain. However the ortholog of these genes identified in *T. spiralis* is significantly closer to the *Hox3* genes from other phyla with two extra characteristic residues in the homeodomain and fewer residues divergent from other Hox genes. Thus identifying the less derived sequence from *T. spiralis* confirms the identity of these genes as divergent members of the *Hox3* group. Again it is apparent that these genes have evolved rapidly away from their orthologs in other phyla. The fact that the *T. spiralis* gene is less derived indicates that many of these changes have occurred within the nematode lineage. Some residues are conservative changes from *Hox3* characteristic residues to residues charcacteristic of the nematode *Hox3* group, such as the change from methionine to leucine at position 34 of the homeodomain (Fig 2.3.). As the *C. elegans* genome does not contain any ortholog of this group they have been called collectively *hox-3* genes prefixed by their species abbreviation maintaining the naming convention for nematode genes and indicating their Hox orthology assignment in line with other phyla.

2.4.3. Deformed/Hox4/lin-39 orthology group genes in nematodes

Their has been some doubt whether the *C. elegans lin-39* belonged to the *Dfd/Hox4* orthology group or not, although most recently it has been referred to as being the *C. elegans Dfd* ortholog (Grandien and Sommer, 2001). The problem with this assignment was that while the *C. elegans* gene has 77% identity to *Dfd*, *Dfd* itself is much closer to Antp (82% identity) and the other central genes. In addition the *lin-39* gene contains an intron which is present in labial and *pb* family genes but absent from the *Dfd* family. However as we have already seen all of the *C. elegans* genes

with the exception of *php-3*, are distant from their orthologs in other phyla. Cloning of a number of *lin-39* orthologs from other nematode species supports the assignment of *lin-39* to the *Dfd* orthology group (Fig 2.4.). However, the characteristic *Dfd* peptide motif found in Dfd genes from other phyla does not appear in the Nematoda ((de Rosa et al., 1999),see chapter 3).

Again this group of nematode genes has a number of characteristic residues within the homeodomain and they have most likely arisen within the nematode lineage, as the ortholog from *T. spiralis* does not share all of them. In two cases instead of the characteristic nematode residue *T. spiralis* has maintained the ancestral residue. In addition the *lin-39* ortholog from this species has maintained an aspartic acid residue characteristic of the *Dfd* group absent in the other nematode species.

It appears that the evolution of the *lin-39* orthologs has been much slower once the major clades III-V have radiated than the two previous orthology groups described. In addition two *lin-39* genes were cloned form *M. javanica*. This gene duplication probably occurred recently (at least since the divergence of the *M. javanica* lineage from that of *S. ratti*, its closest relative in this study) as both genes have shared derived residues in the homeodomain not present in the other species.

Fig 2.3. Alignment of *Hox3* genes

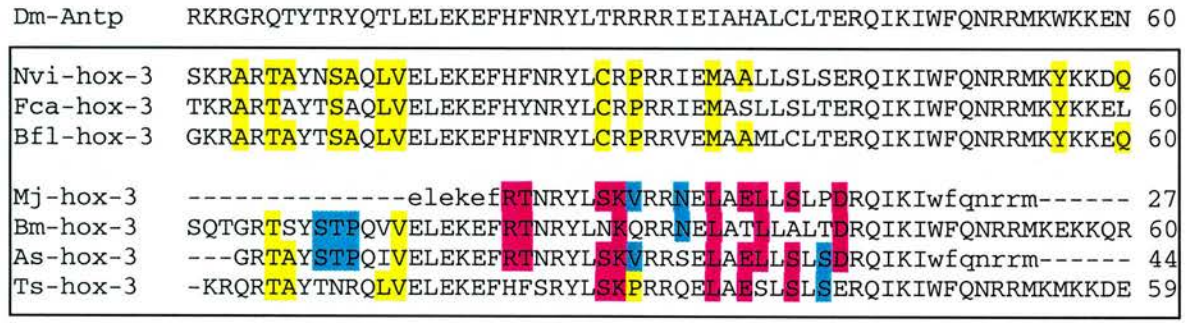


Fig 2.3. Alignment of *Hox3* genes

An alignment of the nematode *Hox-3* genes with *Hox-3* from other phyla compared with the *Drosophila melanogaster Antennapaedia* homeodomain. Yellow residues indicate characteristic residues for the orthology group defined as being different from the *Antp* homeodomain and present in more than 50% of known orthologs (after de Rosa et al 1999). Purple residues represent those likely to be characteristic of the nematode group by the same criteria for characteristic residues across phyla, different from both the *Antp* homeodomain and any characteristic residues for the orthology group. Turquoise residues are those different from the *Antp* homeodomain and any characteristic residues for the orthology group but shared by more than one nematode species. Numbers at the end of each line indicate the number of known residues in the homeodomain.

Species abbreviations are Deuterostomata: Bfl, *Branchiostoma floridae*. Ecdysozoa: Fca, *Folsomia candida*; Dm, *Drosophila melanogaster*. Lophotrocozoa: Nvi, *Nereis virens*. Nematoda: Mj, *Meloidogyne javanica*; Bm, *Brugia malayi*; As, *Ascaris suum*; Ts, *Trichinella spiralis*

2.4.4. Central Hox cluster genes from nematodes

The only central class Hox gene in *C. elegans* the *mab-5* gene originally classed as the homolog of all the central class genes in *Drosophila maelanogaster* (Burglin and Ruvkun, 1993; Burglin et al., 1991). However, if nematodes are part of the Ecdysozoa then the other central genes must have been lost at some point in the *C. elegans* lineage. Homologs of *mab-5* have been found throughout the nematode lineage, except for in the clade I species *T. spiralis* where neither of the central genes is obviously a *mab-5* ortholog (Fig 2.5). The *mab-5* orthology group is similar to the *lin-39* group in that a number of characteristic residues arose in the homeodomain before the radiation of the major clades III-V. Where these characteristic residues have changed they agree with the current view of nematode phylogeny, as members of the same major clade nearly always share them. The orthology relationship of the *mab-5* group genes to Hox genes from other phyla remains unclear. The gene is most likely an ortholog of either the *Lox-5/Hox6-8/ftz* group or the *Antp* group. Previous work has favoured a grouping with the *Lox5//ftz group* (Telford, 2000a).

Apart from the *mab-5* group five other nematode genes that are likely to belong to central orthology groups were cloned. Both *B. malayi* and *A. suum* have a central class gene most like the *Lox5/Hox6-8/ftx* or *Antp* central class genes (Fig 2.5). They have been provisionally called *Bm-ant-1* and *As-ant-1* respectively as they are the first Antennapaedia class genes identified in each species, absent from *C. elegans*, and of uncertain orthology. These genes share derived residues indicating that they are orthologs of each other. They are both closer in sequence to their orthologs in other phyla than the *mab-5* genes are. At this stage, without a more complete phylogenetic analysis these two genes will be considered as a separate orthology group within the Nematoda representing either the *Lox5/Hox6-8/ftx* or *Antp* class gene. An alternate possibility is that these genes are recent paralogs of the *mab-5* group having arisen through gene duplication, for example, within the nematode lineage.

Two central Hox cluster genes closest to either *Lox5/Hox6-8/ftx* or *Antp* orthology groups were also identified from *T. spiralis*. These genes have been called *Ts-ant-1* and *Ts-ant-3* genes respectively as they were the first and third Hox genes

found in *T. spiralis* without obvious *C. elegans* homologs. At this point each gene had been assigned to either the *mab-5* group or as a possible ortholog of the *Bm-ant-1/As-ant-1* group. The *Ts-ant-3* is closer to the *mab-5* group and will be considered as belonging to this group unless further analysis proves otherwise. *Ts-ant-1* has been placed with the genes from *B. malayi* and *A. suum* as it is closer to these than to *mab-5*. A third gene from *T. spiralis*, *Ts-ant-2*, has a very derived homeodomain sequence and cannot be placed in any orthology group.

One central class gene other than a *mab-5* ortholog was also cloned from *S. ratti* and has been called *Sr-ant-1*. This gene is closest to the *Lox2/Ubx* orthology group present in both the Lophotrochozoa and the Ecdysozoa. An ancestral gene representing the *Lox2/Ubx/Lox4/abd-A* orthology groups are thought to have been present in the common ancestor of these two major protostome clades, rather than have arisen by independent gene duplications in each lineage. Although it remains unclear whether the subsequent duplication, giving rise to the proposed abd-A/Lox4 orthology group, occurred before or after these lineages diverged (although the most parsimonius scenario is the duplication occurred before the lineages diverged, (Telford, 2000c)). The short sequence from the nematode is more like *Lox2* than *Ubx*. The nematode gene has five residues in common with *Lox2* and not *Ubx* across the known sequence (Fig 2.6). Previous work has established that the *Lox2* group is characteristic of the Lophotrochozoa (de Rosa et al., 1999). The identification of a possible *Lox2* family gene in the Nematoda would refute this. Ideally a full homeodomain and flanking residues from this gene would help to provide stronger support for this surprising finding, the implications of which are discussed in more detail later. Until such data is available though we must remain sceptical about whether this gene is a real nematode gene or more likely a contaminant from elsewhere, even though to my knowledge no lophotrochozoans have been in the laboratory in which *Sr-ant*-1 was cloned

Fig 2.4. Alignment of *lin-39/Dfd* genes

Fig 2.4. Alignment of *lin-39/Dfd* genes

An alignment of the nematode *lin-39* genes with *Dfdl/Hox4* genes from other phyla compared with the *Drosophila melanogaster Antennapaedia* homeodomain. Yellow residues indicate characteristic residues for the orthology group defined as being different from the *Antp* homeodomain and present in more than 50% of known orthologs (after de Rosa et al 1999). Purple residues represent those likely to be characteristic of the nematode group by the same criteria for characteristic residues across phyla, different from both the *Antp* homeodomain and any characteristic residues for the orthology group. Turquoise residues are those different from the *Antp* homeodomain and any characteristic residues for the orthology group but shared by more than one nematode species. Numbers at the end of each line indicate the number of known residues in the homeodomain. Translations of the primer sequences used to amplify fragments are written in lower case when a fragment has not been extended by an alternative method.

Species abbreviations are Deuterostomata: Bfl, *Branchiostoma floridae*. Ecdysozoa: Dm, *Drosophila melanogaster*. Lophotrocozoa: Nvi, *Nereis virens*. Nematoda: Ce, *Caenorhabditis elegans*; Pp, *Pristionchus pacificus*; Sr, *Strongyloides ratti*; Mj, *Meloidogyne javanica*; Bm, *Brugia malayi*; As, *Ascaris suum*; Ts, *Trichinella spiralis*.

.Fig 2.5. Alignment of nematode central hox genes genes

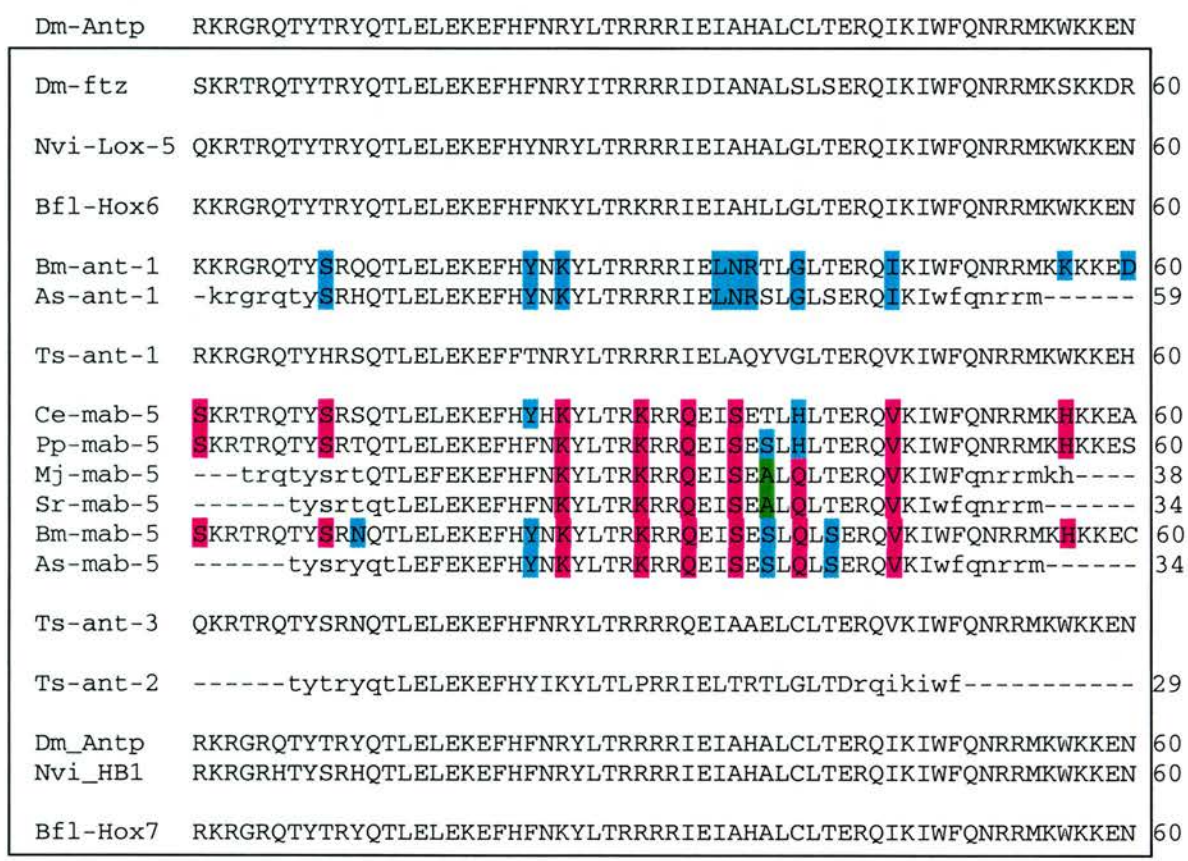


Fig 2.5. Alignment of central hox genes genes

An alignment of the nematode central Hox genes with *ftzl/Lox5/Hox6-8* and *Antp* genes from other phyla. Purple residues represent those likely to be characteristic of the nematode group by the same criteria for characteristic residues across phyla, different from both the *Antp* homeodomain and any characteristic residues for the orthology group. Turquoise and green residues are those different from the *Antp* homeodomain and any characteristic residues for the orthology group but shared by more than one nematode species. The *Ts-ant-3* and *Ts-ant-1* genes have been assigned to the *mab-5* group and the group represented by *Bm-ant-1* and *As-ant-1* respectively (see chapter 4). Numbers at the end of each line indicate the number of known residues in the homeodomain. Translations of the primer sequences used to amplify fragments are written in lower case when a fragment has not been extended by an alternative method.
Species abbreviations are Deuterostomata: Bfl, *Branchiostoma floridae*. Ecdysozoa: Dm, *Drosophila melanogaster*. Lophotrocozoa: Nvi, *Nereis virens*. Nematoda: Ce, *Caenorhabditis elegans*; Pp, *Pristionchus pacificus*; Sr, *Strongyloides ratti*; Mj, *Meloidogyne javanica*; Bm, *Brugia malayi*; As, *Ascaris suum*; Ts, *Trichinella spiralis*.

2.4.5. Posterior group genes in the Nematoda

C. elegans has three posterior group genes in the fully sequenced genome. *egl-5* and *nob-1* appear to be very derived making their orthology assignment unclear while *php-3* is an ortholog of the *AbdB* posterior group found in other ecdysozoan phyla (de Rosa et al., 1999; Ruvkun and Hobert, 1998; Van Auken et al., 2000). The presence of *php-3* in the *C. elegans* genome was discovered very late in the genome sequencing effort and provides the only Hox gene data that supports the inclusion of the nematodes within the Ecdysozoa. An alternate hypothesis is that the *AbdB* Hox signature represents the ancestral protostome state, possibly absent from the members of the Lophotrochozoa sampled so far.

Orthologs of the *egl-5* gene have been found throughout the phylum Nematoda and are relatively well conserved like the *lin-39* and *mab-5* groups (Fig 2.7). In this case *T. spiralis* contains an obvious close ortholog of the *egl-5* group. This confirms the *egl-5* gene was present with all its derived residues at the root of the nematode lineage rather than the alternate hypothesis that its derived homeodomain sequence has evolved at some point during the evolution of the different nematode lineages. Interestingly three of the four residues within the homeodomain likely to be mostly responsible for the inclusion of *php-3* in the *AbdB* group are also conserved in the *egl-5* homeodomain. It is therefore possible that the *egl-5* group represents a paralogous gene of the *AbdB* group that arose at the root of the Nematoda and underwent rapid evolution away from the *AbdB* sequence while maintaining the characteristic residues. These genes have all been named after the their ortholog in *C. elegans*.

Orthologs of the *C. elegans php-3* gene have also been found throughout the Nematoda and are relatively highly conserved (Fig 2.8). They all contain characteristic residues for posterior group genes as well as those that drive their inclusion within an *AbdB* posterior gene group specific to the Ecdysozoa. There are also two residues that are characteristic of the nematode *php-3* group genes shared by all the species. These genes have also been named according to the current convention after their *C. elegans* ortholog. Of all the nematode Hox gene groups the

php-3 family is the closest to its orthologs in other phyla (with the possible exception of the possible *Lox2/Ubx* gene found in *S. ratti*).

The third *C. elegans* posterior gene, *nob-1*, was not found in any of the other nematode species in this study. The derived sequence of this gene may explain why it hasn't been identified by the degenerate PCR approach used here. Alternatively it may be the result of more recent gene duplication within the *C. elegans* genome, which is absent from other nematodes. In support of this a large genomic clone containing the *Pp-php*-3 gene has been identified and a *nob*-1 ortholog has not been identified. So either *nob-1* is absent from *P. pacificus* or it is not in close proximity to *php-3* as it is in the *C. elegans* genome (Ralf Sommer, personal communication). However, only complete genome sequencing of another nematode will confirm either possibility.

Fig 2.6. Alignment of a nematode *Ubx/Lox2* gene with those from other phyla

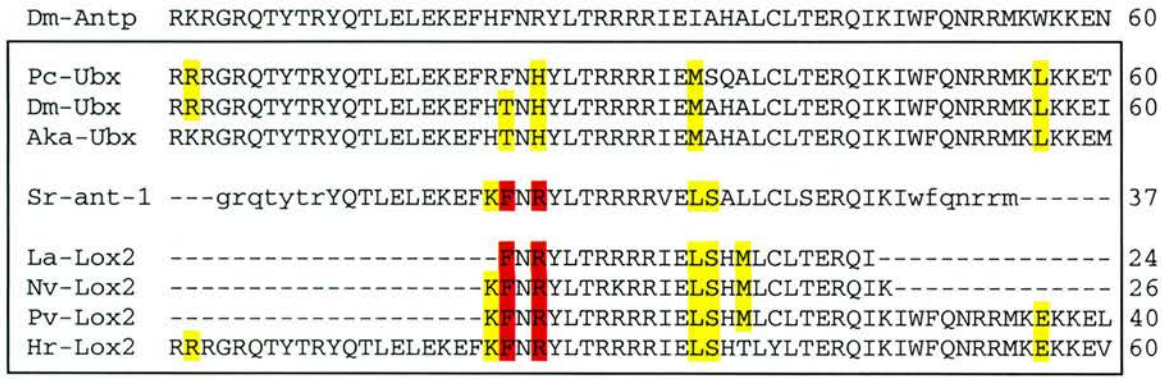


Fig 2.6. Alignment of a nematode *Ubx/Lox2* gene with those from other phyla

An alignment of a nematode *Ubx/Lox2* gene with those from other phyla showing that the nematode gene is more like the *Lox2 group*. Yellow residues are those characteristic for either the Lophotrochozoan *Lox2* gene of the Ecdysozoan *Ubx* gene or both (as defined above). The nematode gene shares characteristic residues with the *Lox2* group but not with the *Ubx* group. Red residues, although not characteristic, support the grouping of *Sr-ant-1* with *Lox2* and not *Ubx*. Numbers at the end of each line indicate the number of known residues in the homeodomain. Translations of the primer sequences used to amplify fragments are written in lower case when a fragment has not been extended by an alternative method.

Species abbreviations are Ecdysozoa: Dm, *Drosophila melanogaster*; Pc, *Priapulus caudatus*; Aka, *Acanthocara kaputensis*. Lophotrocozoa: La, *Lingula anatina*; Nv, *Nereis virens*; Pv, *Patella vulgata*; Hr, *Helobdella robusta*. Nematoda: Sr, *Strongyloides ratti*.

Fig 2.7. Alignment of the egl-5 group with posterior genes from other phyla

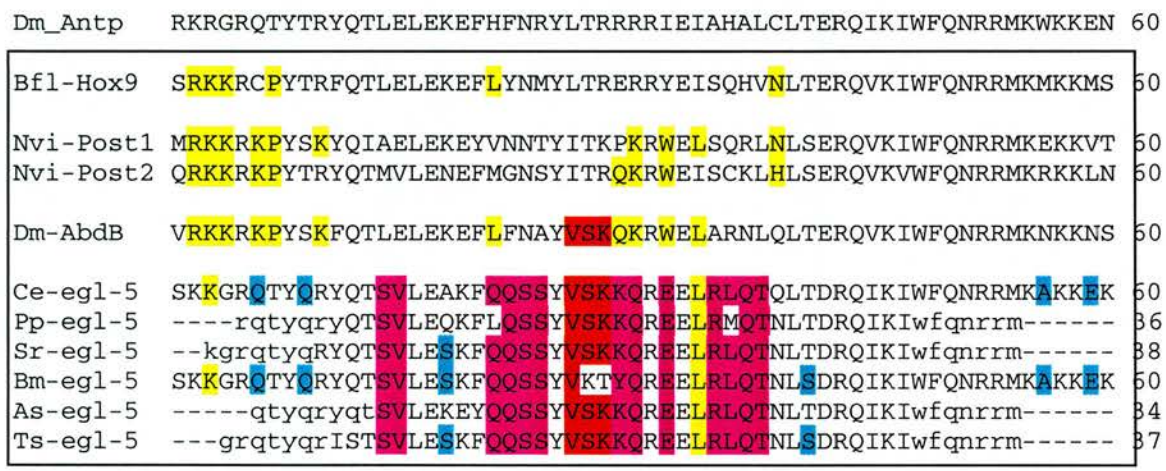


Fig 2.7. Allignment of the egl-5 group with posterior genes from other phyla

An alignment of the nematode *egl-5* genes with posterior genes from other phyla. Yellow residues indicate characteristic residues for the orthology group defined as being different from the *Antp* homeodomain and present in more than 50% of known orthologs (after de Rosa et al 1999). Purple residues represent those likely to be characteristic of the nematode group by the same criteria for characteristic residues across phyla, different from both the *Antp* homeodomain and any characteristic residues for the orthology group. Turquoise residues are those different from the *Antp* homeodomain and any characteristic residues for the orthology group but shared by more than one nematode species. Red residues are those characteristic of the Ecdysozoa posterior *AbdB* group shared by the *egl-5* group. Numbers at the end of each line indicate the number of known residues in the homeodomain. Translations of the primer sequences used to amplify fragments are written in lower case when a fragment has not been extended by an alternative method.
Species abbreviations are Deuterostomata: Bfl, *Branchiostoma floridae*. Ecdysozoa: Dm, *Drosphila melanogaster*. Lophotrocozoa: Nvi, *Nereis virens*. Nematoda: Ce, *Caenorhabditis elegans*; Pp, *Pristionchus pacificus*; Sr, *Strongyloides ratti*; Mj, *Meloidegyne javanica*; Bm, *Brugia malayi*; As, *Ascaris suum*; Ts, *Trichinella spiralis*.

Fig 2.8. Alignment of the *php-3* group with posterior genes from other phyla

Fig 2.8. Allignment of the *php-3* group with posterior genes from other phyla

An alignment of the nematode *php-3* genes with posterior genes from other phyla compared with the *Drosophila melanogaster Antennapaedia* homeodomain. Yellow residues indicate characteristic residues for the orthology group defined as being different from the *Antp* homeodomain and present in more than 50% of known orthologs (after de Rosa et al 1999). Purple residues represent those likely to be characteristic of the nematode group by the same criteria for characteristic residues across phyla, different from both the *Antp* homeodomain and any characteristic residues for the orthology group. Turquoise residues are those different from the *Antp* homeodomain and any characteristic residues for the orthology group but shared by more than one nematode species. Red residues are those characteristic of the Ecdysozoa posterior *AbdB* group shared by the *php-3* group. Numbers at the end of each line indicate the number of known residues in the homeodomain. Translations of the primer sequences used to amplify fragments are written in lower case when a fragment has not been extended by an alternative method.

Species abbreviations are Deuterostomata: Bfl, *Branchiostoma floridae*. Ecdysozoa: Dm, *Drosophila melanogaster*. Lophotrocozoa: Nvi, *Nereis virens*. Nematoda: Ce, *Caenorhabditis elegans*; Pp, *Pristionchus pacificus*; Sr, *Strongyloides ratti*; Bm, *Brugia malayi*; As, *Ascaris suum*; Ts, *Trichinella spirali*

2.5. Hox gene loss within the phylum Nematoda

The identification of at least 3 new Hox gene orthology groups within the phylum Nematoda proves that Hox gene loss has occurred within the phylum. As the *C. elegans* genome is the only one that has been completely sequenced gene loss can only be inferred in this species. However we can make statements about the earliest possible time at which particular Hox genes were lost from the lineage leading to *C. elegans*. The earliest time at which both the *Hox3* and *Ubx/Lox-2* orthology groups could have been lost is just after the division between clades IV and V (Fig 2.9) By the same logic the earliest point at which either the Lox5/Hox6-8/ftz or *Antp* orthology group (depending which one *mab-5* represents) could have been lost is at the division between clade III and clades IV and V (Fig 2.9).

The possibility that *Sr-ant-1* might be closer to the *Lox2* genes of the Lophotrochozoa suggests that this gene is not restricted to this lineage. If it is a true *Lox2* ortholog than it suggests that the *Ubx* gene, far from being characteristic of the Ecdysozoa may actually be characteristic of the Arthropoda and close relatives. One implication of this is that Priapulids would be closer to the Arthropoda than the Nematoda as they have the *Ubx* gene (de Rosa et al., 1999). If the *Sr-ant-1* gene proves to be a *Lox2* like gene when further sequence is cloned it will be an example of the danger of using Hox genes as characteristic signatures of evolutionary lineages before a wide phylogenetic sampling is available (Kobayashi et al., 1999; Telford, 2000c). However, until the authenticity of *Sr-ant-1* is proved beyond doubt by cloning the rest of the coding sequence and it is safer to assume that it is a contaminant.

Three other Hox orthology groups that should have been present before the divergence of the nematode lineage from the last common ancestor of the Ecdysozoa have not yet been identified in nematodes, and this is therefore the earliest point at which gene loss could have occurred. Hox gene loss has previously been observed in the mouse genome where Hox orthology groups are missing from one or more other copies of the 4 paralogous clusters (Holland and Garcia-Fernandez, 1996). However, every orthology group remains represented in the mouse genome. Loss of a whole orthology group (*Hox7*) has been observed in the puffer fish *Fugu rugripes* (Aparicio

et al., 1997) and possibly also in barnacles (Mouchel-Vielh et al., 1998). These changes are not on the same scale as described here: loss of more than half the Hox cluster membership accompanied by loss of structural colinearity. Some orthology groups may be absent from the whole phylum Nematoda.

2.5.1. The tempo of Hox gene loss within the Nematoda

Pinpointing exactly when during the evolution of the Ecdysozoa and the Nematoda observed Hox gene losses occurred is almost impossible. Further searching among nematode species may reveal orthologs of the remaining missing groups within the Nematoda or it may not. Only full genome sequencing of other nematodes will allow more precise statements to be made. However, the missing orthology groups (*pb/Hox2, Scr/Hox5, Ubx/Lox2*) may be found by more directed approaches within the Nematoda. One problem with finding two more central orthology groups is the observation that the central groups genes that have been found in Nematoda, including *lin-39* and *mab-5* appear to have divergent homeodomains. It is possible that any assignment to an orthology group beyond the statement that they are central class genes will not be possible (see chapter 4). Previous authors have used the homology of flanking residues outside of the homeodomain to aid in the assignment to orthology groups (de Rosa et al., 1999; Telford, 2000a). Unfortunately the flanking residues for genes for which this data would be most informative, are either not known (*Sr-ant-1*, for example) or in most cases not informative (see chapter 3 for description). This reflects both a loss of characteristic peptides (this is certainly the case for the *lin-39* group which has lost the *Dfd/Hox4* peptide motif) and the situation that the current sample of flanking residues doesn't include orthology groups with characteristic flanking peptide motifs.

Loss of at least three orthology groups within the phylum since the divergence of clades III, IV and V implies that these genes no longer had essential functions during the evolution of the *C. elegans* lineage. Loss of six orthology groups takes some explanation in terms of the evolution of developmental mechanisms. Without the full sequence of another nematode genome or at least a full

map of a second nematode Hox cluster we can only discuss the different modes and tempos of gene loss that are possible within the current data set.

The first extreme scenario is that all the orthology groups have been lost recently from the *C. elegans* lineage and nematodes outside of clade V, say, have all the orthologous groups characteristic of the other protostomes (if not the Ecdysozoa). If this is the case the current sampling has failed to identify 3 orthology groups altogether and failed to pick many orthology groups from different individual species. This is obviously not impossible especially if the sequences of these missing Hox genes have diverged significantly as appears to be the case for many of the genes that have been found (see Fig 2.9). The discovery of the *Sr-ant-1* gene in clade IV but not of any ortholog in clade III or clade V illustrates this point. The second and probably more likely scenario is that various orthology groups have been lost through the Nematoda with *C. elegans* and its closest relatives having lost six orthology groups since the root of Ecdysozoa. Other nematode clades may have lost somewhere between zero and four orthology groups and presence of an orthology group in a less basal clade does not preclude loss or partial loss of that orthology group from another more basal clade. For example clade IV may have lost the ortholog of the *Bm-ant-1/As-ant-1* gene and clade III may have lost the ortholog of the *Sr-ant-1* gene (if this gene proves not to be a contaminant) since the common ancestor of both clades, which had both genes.

Whatever the exact tempo of gene loss within the Nematoda, the position of the Nematoda within the within the Ecdysozoa suggests that some major change(s) in the patterning of the anteroposterior axis must have occurred to allow the loss of Hox genes. We have no knowledge of the expression patterns of Hox genes outside of *C. elegans* (accept analysis of the mutant phenotypes of *Pp-mab-5* and *Pp-lin-39*). Knowing the relative expression patterns of the orthology groups not found in *C. elegans* will be essential to explain the evolutionary events that have led to Hox gene loss.

2.5.2. What evolutionary changes in the developmental mechanisms of the Nematoda and particularly C. elegans may have led to Hox gene loss?

This question could just have easily been asked before any knowledge about the absence or presence of Hox genes elsewhere in the phylum. However, the new data helps to refine the possible hypotheses, identifies substrates to test them on and provides some potential tools for this. As previously discussed the evolution of precise cell lineage determined mode of development in *C. elegans* (and probably other nematodes as well) could have arisen concomitantly with Hox gene loss. It is also apparent that *C. elegans* has lost other genes central to developmental processes, such as Hedgehog and the ParaHox genes (Ruvkun and Hobert, 1998). The loss of these genes suggests that at some point their functions have become redundant If these two observations are indeed causally linked then the existence of more Hox orthology groups in the other Nematoda might suggest that their developmental mode is not so biased towards a cell-lineage determined mode of development and that the roles of these genes extra Hox genes have not become redundant. However, available data about cell-lineage from other nematodes suggests that it is highly determined and reproducible at least in clades III, IV and V (Voronov et al., 1998), even if constant cell number is not (Cunha et al., 1999). The presence of extra Hox genes in nematodes of these clades suggests that Hox gene loss is not directly related to the adoption of a cell-lineage mode of development.

If the other nematode clades represent intermediate stages of Hox gene loss in comparison to what is observed in *C. elegans* we might expect the expression patterns of the extra Hox genes to give clues as to evolution of developmental processes that have led to their loss from *C. elegans*. Do their expression patterns correspond to areas of the body axis missing from *C. elegans*? Considering the conserved body plan of the Nematoda this seems unlikely. More likely the extra Hox genes will have functions performed by other genes in *C. elegans,* genes that have adopted this role perhaps as a result of the cell lineage based mode of development obviating the need for Hox gene patterning. Alternatively extra Hox genes may have evolved alternative roles in other nematode lineages, perhaps like the apparently post-embryonic roles adopted by *lin-39*, *mab-5* and *egl-5*. In fact triple mutant

animals for these genes can be viable discounting overlapping but redundant embryonic functioning of these three more centrally located Hox genes (Van Auken et al., 2000; Wrischnik and Kenyon, 1997). As previously discussed only the anterior Hox gene *ceh-13* and posterior Hox genes *php-3* and *nob-1* have embryonic pattern functions and are required for correct embryogenesis.

Fig 2.9. Overview of Hox gene evolution through the phylum Nematoda

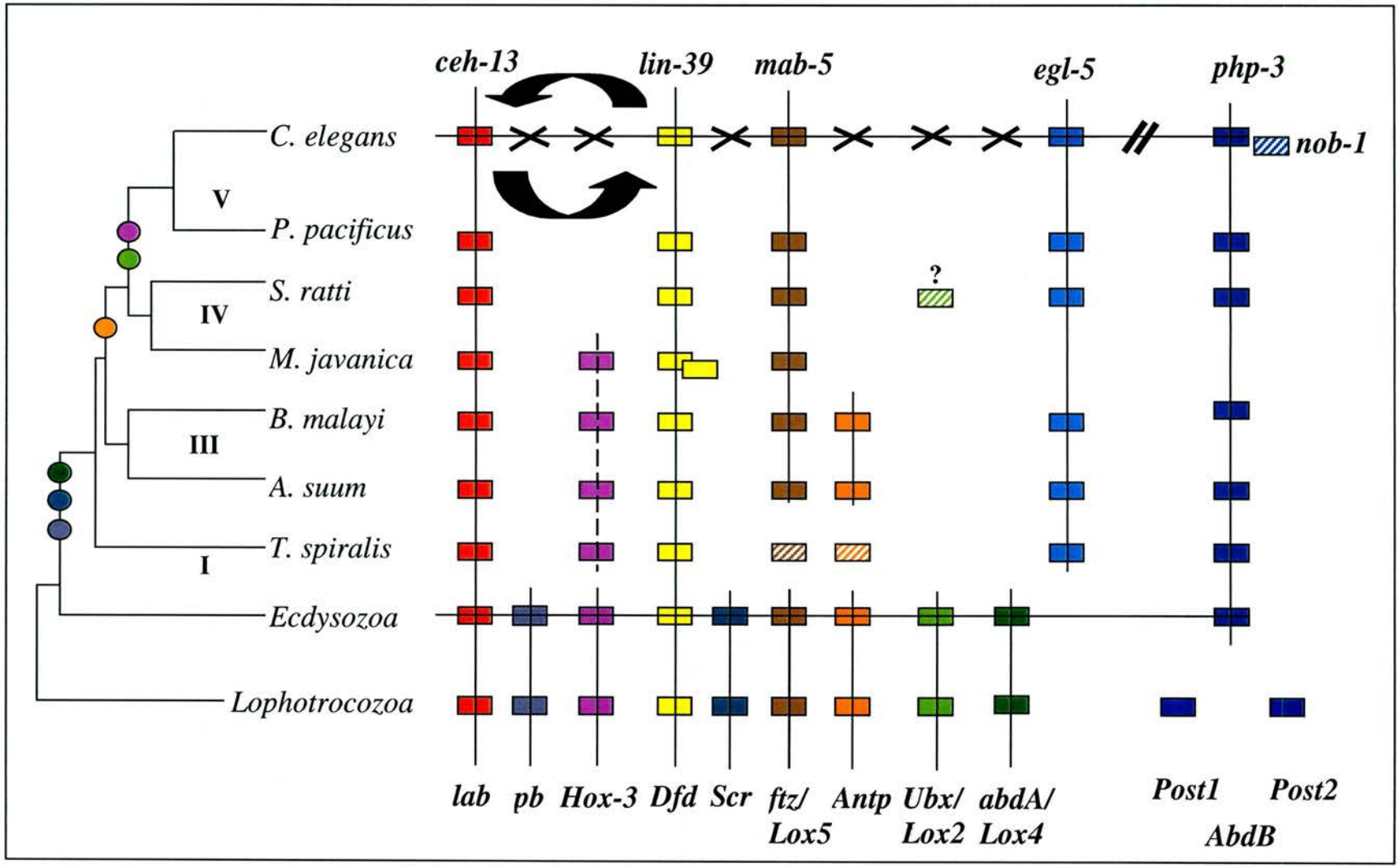


Fig 2.9. Overview of Hox gene evolution through the phylum Nematoda

An overview of the Hox genes identified form the phylum Nematoda. Boxes represent the presence of Hox genes within a species and are colour coded to represent different orthology grouped together vertically and arranged according to colinearirty horizontally. The orthology groups are given abbreviated names for the nematodes at the top of the columns and the corresponding names for the Ecdysozoa and where appropriate the Lophotrochozoa at the bottom. Horizontal lines indicate that genes are linked (although loosely in *C. elegans*). Vertical lines indicate that orthology relationships are certain as opposed to proposed (dashed lines) or uncertain (no lines). The curled arrows indicate the inversion of *ceh-13* and *lin-39* on chromosome 3 pf *C. elegans*. Boxes filled with striped colours indicate genes for which orthology assignment remains uncertain. These assignments are tested using phylogenetic analysis in chapter 5. Black crosses in *C. elegans* indicate genes that are assumed to have been lost at some point during the evolution of this lineage as they are absent from the complete genome sequence. A question mark above a box indicates a gene that does not fit the phylogenetic pattern outlined by other authors (de Rosa et al., 1999).

The phylogenetic tree to the left represents the relationship of the nematode species with each other and the rest of the protostomes split into the Ecdysozoa and Lophotrochozoa (Blaxter et al., 1998; de Rosa et al., 1999). Coloured circles on the tree are colour coded to match the orthology groups of Hox genes missing from *C. elegans*. Their position on the tree represents the earliest time in the evolution of the lineage leading to *C elegans* that the genes could have been lost based on the data represented in the main part of the figure. Thus at least 3 Hox gene losses have occurred within the phylum and these have all occurred since the division of clades IV and V from the other nematode lineages. The presence or absence of genes not identified in this screen from the other species of nematode cannot be commented on until the full genomes of these species are sequenced.

The evolution of Hox genes in nematodes has already proved a useful substrate for studying the evolution of developmental mechanisms, and the overview of Hox genes across the phylum strongly suggests that it will continue to do so. Some understanding of the processes involved in allowing the loss of highly conserved genes will undoubtedly greatly complement our understanding of the evolution of gene function and the interactions between intertwined genetic networks. Studying Hox genes outside of the free-living nematodes of clade V will be difficult but the use of basic and the development of more advanced molecular genetic techniques is not impossible (see chapters 3 and 6 of this thesis). In addition clade IV nematodes include free-living nematodes that that have potential for use as satellite models in a similar manner to *P. pacificus* (Felix et al., 2000). If Hox genes absent from *C. elegans* can be identified from these species and the power of basic genetic and molecular analysis brought to bear then it should eventually be possible to investigate the function of the nematode Hox genes absent from *C. elegans* (Aboobaker, wish list for 2002).

2.6.Acknowledgements

The work described in this chapter linked to the further discussions and conclusions in chapter 4 were begun over 3 years ago as a ten week rotation project in the laboratory of Dr Mark Blaxter. Somehow everything in this chapter is irrevocably attached to that period of time. Typically Mark was the first to take affirmative action on the possible significance of cloning Hox genes from a nematode only distantly related to *C. elegans*. He also showed great patience with my initial lack of progress. Two other people had the great patience required for the task of helping me to develop green fingers, at the time (and to this day) motivated by very different forces. I am very grateful to both Ms Jennifer Daub and Mr David Guiliano for their time, encouragement and help. I am also very grateful to a very gifted summer student Ms Emma Deas for her hard work and dedication as a Wellcome Trust Vacation Scholar in our laboratory. I would also like to thank all the individuals who provided nematode material.

CHAPTER 3: FURTHER ANALYIS OF HOX GENES FROM NEMATODES.

3.1. Further cloning and characterisation of *B. malayi* Hox genes

The Hox gene sequences generated by the degenerate PCR screen described in the previous study were extended by 5' and 3' RACE experiments. An alternative method for cloning regions to the 5' of the identified sequence was to make use of the SL1 splice leader sequence trans-spliced to the majority of nematode mRNA species (Blaxter and Liu, 1996). The use of a forward primer of the same 22 base pairs (bp) sequence of the spliced leader and a gene-specific reverse primer can allow the rapid identification of 5' ends. A second technique for extending known sequence was to use inverse PCR of genomic DNA preparations. Both these approaches are described in detail in the Materials and Methods section of this thesis. These approaches have led to the cloning the full homeodomains of all except one of the *B. malayi* Hox genes (*Bm-ceh-13*). All nucleic sequences generated in this thesis are presented in Appendix 1 even though only some of them are presented in their entirety in the text. The studies described here lay the foundations for mapping the *B. malayi* Hox cluster, studying the expression patterns of the *B. malayi* Hox genes and performing a comparative molecular analysis of the evolution of these genes with *C. elegans*.

3.2. Cloning of *B. malayi* Hox gene cDNAs and genomic clones

One of the goals for cloning the Hox gene complement of *B. malayi* was to produce a set of reagents for mapping the Hox cluster as part of the ongoing WHO coordinated Filarial Genome Project (Blaxter et al., 1999). Some of the partial homeodomain clones described in the last chapter have been extended and genomic clones spanning intronic regions have been generated. The first steps have been taken in using these as probes to map the Hox cluster of *B. malayi*. In addition the cDNA clones have allowed a preliminary analysis of the expression of Hox genes through the parasite lifecycle. More detailed analysis of expression patterns, and possibly of function (see

chapter 5) and comparative studies with both *C. elegans* and *P. pacificus* will arise from these preliminary efforts.

3.2.1. More than one transcript may arise from the Bm-mab-5 gene

A full length *Bm-mab-5* cDNA was cloned by using 5' and 3' RACE and was found to have the SL1 spliced leader sequence at the 5' terminus of all clones sequenced. The full-length cDNA is 652 bp long including the SL1 leader sequence and encodes a 194 amino acid protein (Fig 3.1). The protein has regions of identity to the Mab-5 proteins of *P. pacificus* and *C. elegans* outside of the homeodomain and the hexapeptide motif (Fig 3.1). These regions of homology are intriguing as they suggest that motifs other than the homeodomain and hexapeptide are conserved and therefore may be important for function (see below). Further analysis of the *Bm-mab-5* gene led to the discovery of an alternate splice site in the gene. On amplifying the *Bm-mab-5* gene from first strand cDNA prepared from pre-blood feed Mf using the SL1 forward primer and a gene specific reverse primer a product of the predicted size was generated. However, a second smaller product was also generated which when cloned and sequenced was found to be a truncated *Bm-mab-5* cDNA with the SL1 leader sequence at the 5' end (Fig 3.1). This putative alternate transcript could be translated with the first possible start codon being the methionine residue of the hexapeptide motif.

Primers designed to the cDNA sequence were used for PCR from gDNA. Three overlapping PCR products were generated which spanned the full length of the *Bm-mab-5* gene. Cloning of the three products proved difficult as screening of positive clones revealed that the inserts of two of the three PCR products were unstable in *E. coli*, apparently unsergoing deletion during prepartaion. The smallest of the three products representing the C-terminus of the gene was approximately 1.8kb. Partial sequencing of this clone revealed an intron in the homeodomain of the *Bm-mab-5* gene, at a position conserved in the *mab-5* genes of *P. pacificus* and *C. elegans*. Another of the PCR fragments, representing the N-terminus of the gene was approximately 3.4kb in length was eventually cloned into a large insert plasmid and was also partially sequenced. This analysis revealed a 3.1kb intron at the position of

the internal splice site described above. Intriguingly this reveals the possibility that there is a promoter region within this intron that may be responsible for the initiation of this alternate transcript. Further sequencing of the intronic region upstream of the internal SL1 site should reveal whether or not this is a possibility. An alternative possibility is that the resolution of the *Bm-mab-5* primary transcription leads to the addition of the SL1 RNA to the terminus of the second exon during resolution of some primary transcripts. However, attempts to amplify this shorter transcript from any stage other than microfilaria has so far failed suggesting that it is stage specific. The final PCR product has still escaped attempts to clone it. Direct sequencing of PCR product revealed the position of the remaining intron in the *Bm-mab-5* gene of approximately 6.8 kb in length. The minimal sequence information available from the introns of this gene indicates that they are very AT rich, consisting of runs of A and T residues.

Genomic clones of *Bm-mab-5* containing the first and third introns of the gene were used to generate probes for chemiluminescent hybridisation, by labelling during PCR with biotin labelled primers. Screening of two different gridded BAC libraries representing an estimated nine fold genome coverage has failed to identify a BAC clone containing the *Bm-mab-5* gene. This lack of success may be related to the instability of regions of the *Bm-mab-5* gene in bacterial cloning vectors described above. This instability may in turn be due to the repetitive A/T rich nature of the intronic regions of this gene. Currently alternative phage genomic libraries and genome walking approaches are being used to identify the region upstream of the *Bm-mab-5* gene in the hope of cloning a putative promoter region. If his region can be cloned and characterised in addition to obtaining the full sequence of the second intron of this gene it should be possible to see if the promoter functions in *C. elegans* by transgenesis. This approach should be applicable to all the *B. malayi* Hox genes, including those that are absent from the *C. elegans* genome.

Fig 3.1 Full length transcript of the *Bm-mab-5* gene

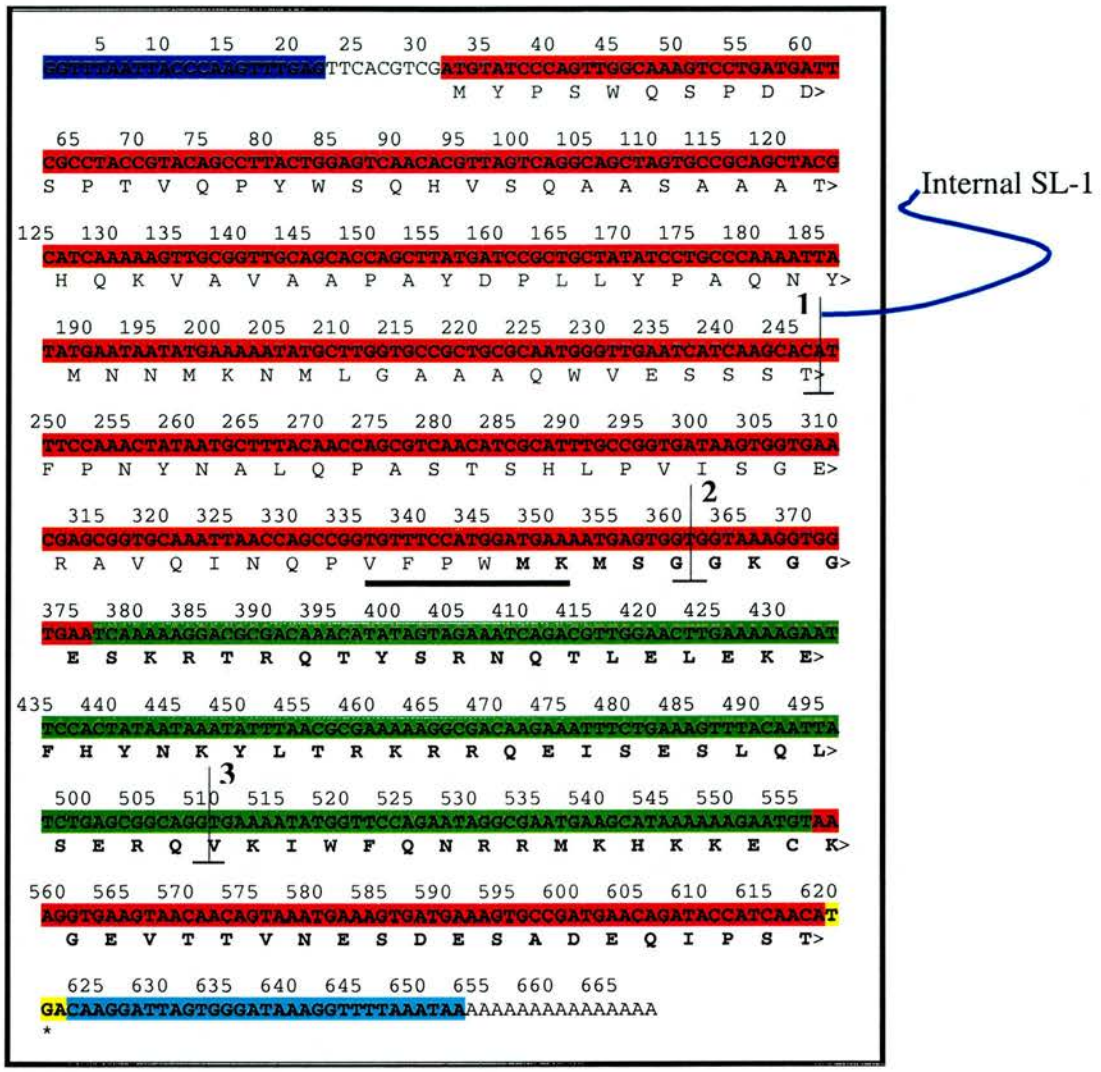


Fig 3.1 Full length transcript of the *Bm-mab-5* gene

A full length *Bm-mab-5* transcript was cloned by 5' and 3' RACE. The region encoding the 60 amino acid homeodomain is highlighted in green, with the rest of the transcript in red, stop codon in yellow and 3' UTR in turquoise. The SL1 leader sequence is highlighted in blue. Introns are indicated by numbers next to vertical lines at the position of the intron. An internal SL1 splice site was identified at the first intron splice sight. The predicted truncated protein produced by this alternate transcript is highlighted in bold and begins within the hexapeptide motif. The hexapeptide motif is underlined by a black bar.

Fig 3.2 Allignment of nematode Mab-5 proteins

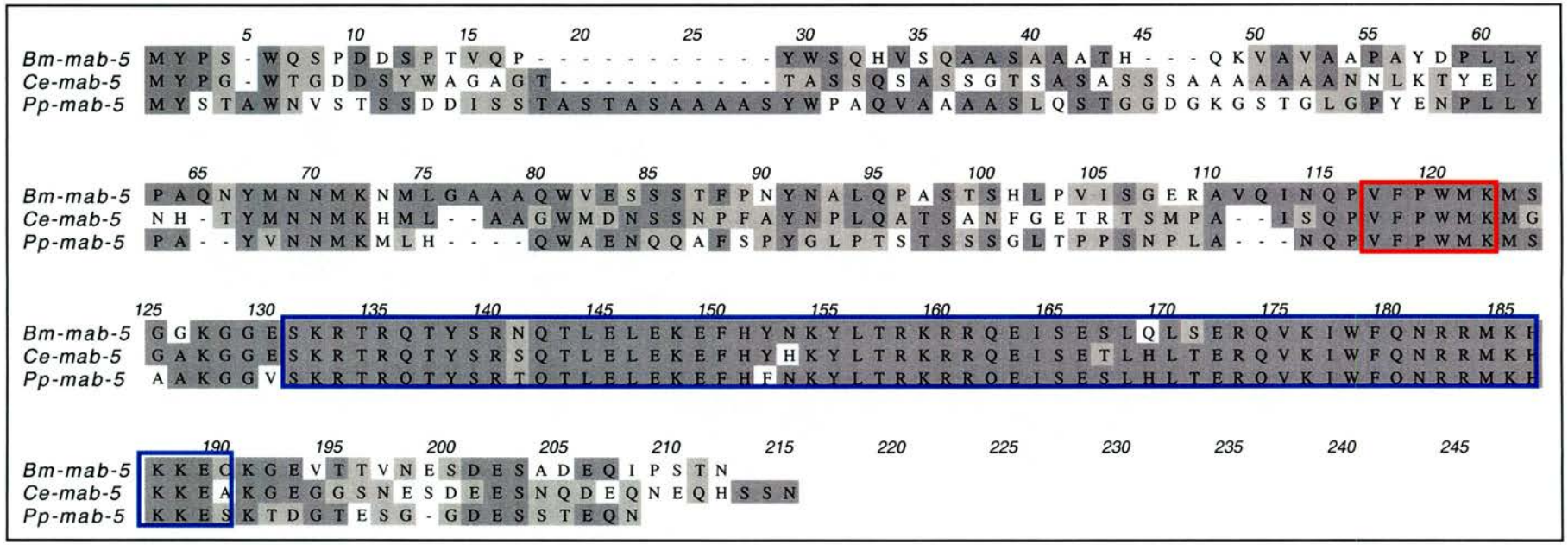


Fig 3.2 Allignment of nematode mab-5 translated products

The MAB-5 proteins of *B. malayi* (Bm), *C. elegans* (Ce) and *P. pacificus* (Pp) are presented. The homeodomain region is boxed in blue and the hexapeptide region is boxed in red. Identities are shaded in dark grey and similarities in light grey. The homeodomain and hexapeptide motif are the only regions characterised as having functional significance. However, the MAB-5 proteins have other conserved motifs and residues outside of these regions. The conservation observed suggests that these regions are likely to be of functional significance.

3.2.2. Comparison of the B. malayi ceh-13 and lin-39 genes to their orthologs in C. elegans and P. pacificus

The 5' ends of both these genes have been successfully cloned using 5' RACE technology. The 3' ends of these transcripts were not identified by a 3' RACE approach, this appears to be due to the presence of lysine rich regions in the C-terminus of the protein the proteins they code for. The method for 3' RACE relies on oligo-dT priming to produce first strand cDNA. It appears that reverse transcription of these transcripts is initiating at the polyA stretches coding for poly lysine and not at the true 3' terminus of the mRNA. Preliminary attempts to extend the partial homeodomain sequences in a 3' direction using inverse PCR have not been fully successsful, although they have extended the original sequences somewhat (see Chapter 2).

The 5' end of the *Bm-ceh-13* transcript is approximately the same length as the *Ce-ceh-13* transcript. This mRNA is also SL1 transpliced like *C. elegans* and *P. pacificus* orthologs. Alignment of Bm-CEH-13 and Ce-CEH-13 reveals regions of homology outside of the homeodomain and hexapeptide (Fig 3.3). As with the alignment of the MAB-5 orthologs this indicates residues that may have functional significance for the protein.

The 5' end of the *Bm-lin-39* gene is significantly shorter than that of *C. elegans* and shorter than that of *P. pacificus* (Fig 3.4). The transcript is also SL1 transpliced with a short 5' untranslated region, consistent with the findings for other *B. malayi* nematode Hox genes. The spacing and sequence of the homeodomain and hexapeptide are conserved but the rest of the N-terminus is not highly conserved and is truncated. Comparative functional analysis of the *P. pacificus* and *C. elegans lin-39* genes has indicated that functional differences between the genes have evolved through changes in regulatory rather than coding sequences. The *Pp-lin-39* gene adequately substitutes for *Ce-lin-39* function when under the control of *C. elegans* regulatory regions (Grandien and Sommer, 2001).

Initial investigation of the genomic structure indicates that *Bm-lin-39* has a conserved intron in the homeodomain also found in both *C. elegans* and *P. pacificus*

Fig 3.3 Alignment of CEH-13 proteins

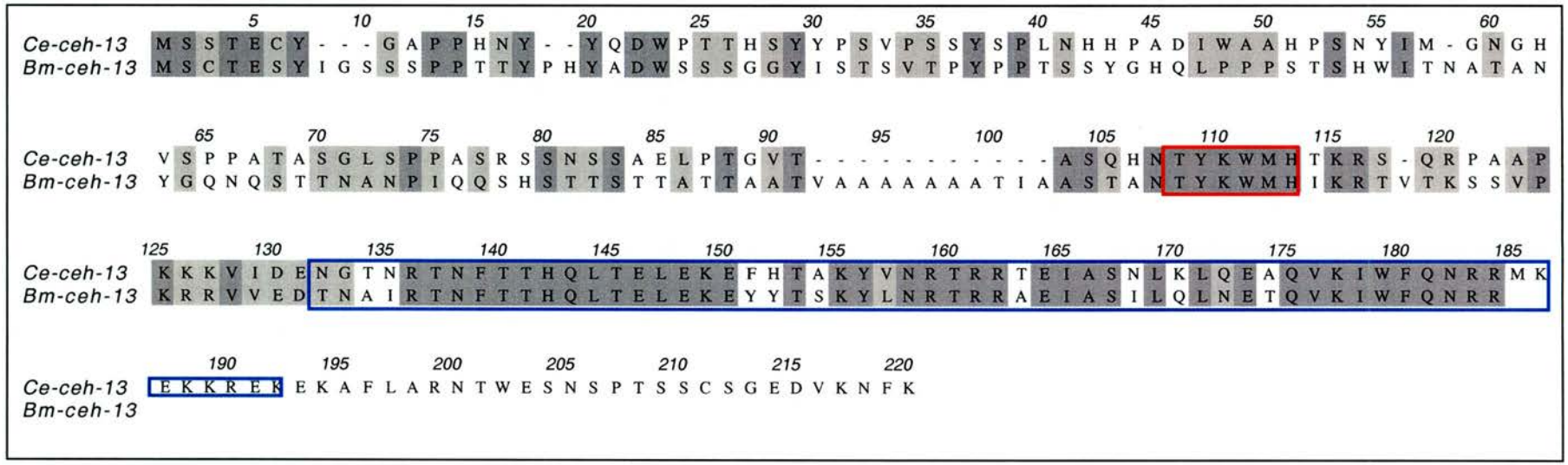


Fig 3.3 Alignment of CEH-13 proteins

An alignment of the CEH-13 proteins of *B. malayi* and *C. elegans*. The homeodomain is highlighted by a blue box and the hexapeptide motif by a red box. Grey shading indicates identities and light grey shading indicates similarities.

Fig 3.4 Alignment of LIN-39 proteins

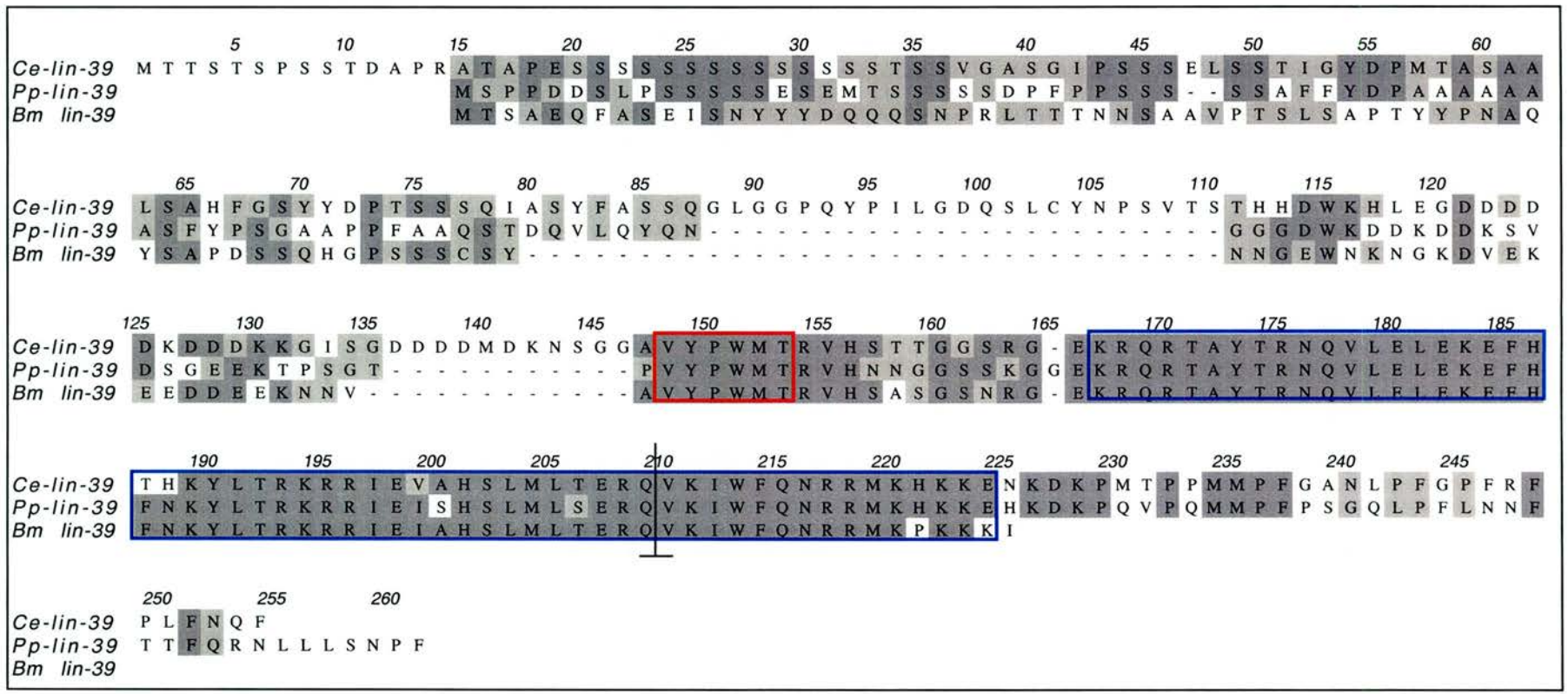


Fig 3.4 Alignment of LIN-39 proteins

An alignment of the LIN-39 proteins of *B. malayi, P. pacificus* and *C. elegans*. The homeodomain is highlighted by a blue box and the hexapeptide motif by a red box. Grey shading indicates identities and light grey shading indicates similarities. The line in the homeodomain indicates the conserved intron position.

The intron is approximately 900 bp and is not as A/T rich as the introns from other *B. malayi* Hox genes. The genomic structure of the 5' end of the gene has not yet been elucidated.

3.2.3. Cloning of the Bm-hox-3 gene and identification of a divergent hexapeptide

The current *Bm-hox-3* cDNA is not full length at its 3' end but includes the full 5' end. The 5' end of the transcript was cloned using 5' RACE and like other nematode Hox genes from Brugia is SL1 transpliced. RT-PCR using the reverse primers located in the homeodomain and the SL1 forward primer did not identify any transcripts of alternative length suggesting that this transcript does not have alternately spliced forms that include the homeodomain.

The 3' end of the gene has been extended by PCR from the *B. malayi* L2 cDNA library in which *Bm-hox-3* expression was detected at significant levels (see below). PCR was performed with a forward primer from the homeodomain and a reverse vector primer. This clone was not full length however as it terminates before the stop codon (see Fig 3.5).

The divergent homeodomain sequence of this gene originally made it difficult to place in a Hox orthology group. Only its homology to the *Ts-hox-3* gene confirmed that it is probably a *Hox3* ortholog. The 5' and 3' end sequence extending beyond the homeodomain has no regions of homology to other Hox genes from other phyla. However a divergent hexapeptide can be discerned (Fig 3.5). The cloning of regions 5' of the homeodomain from other nematode *hox-3* genes may identify regions of homology like those observed for the *mab-5* and *ceh-13* genes. As this gene is missing from *C. elegans* analysis of its expression pattern relative to other Hox genes will be fascinating.

Fig 3.5 A partial *Bm-hox-3* cDNA clone

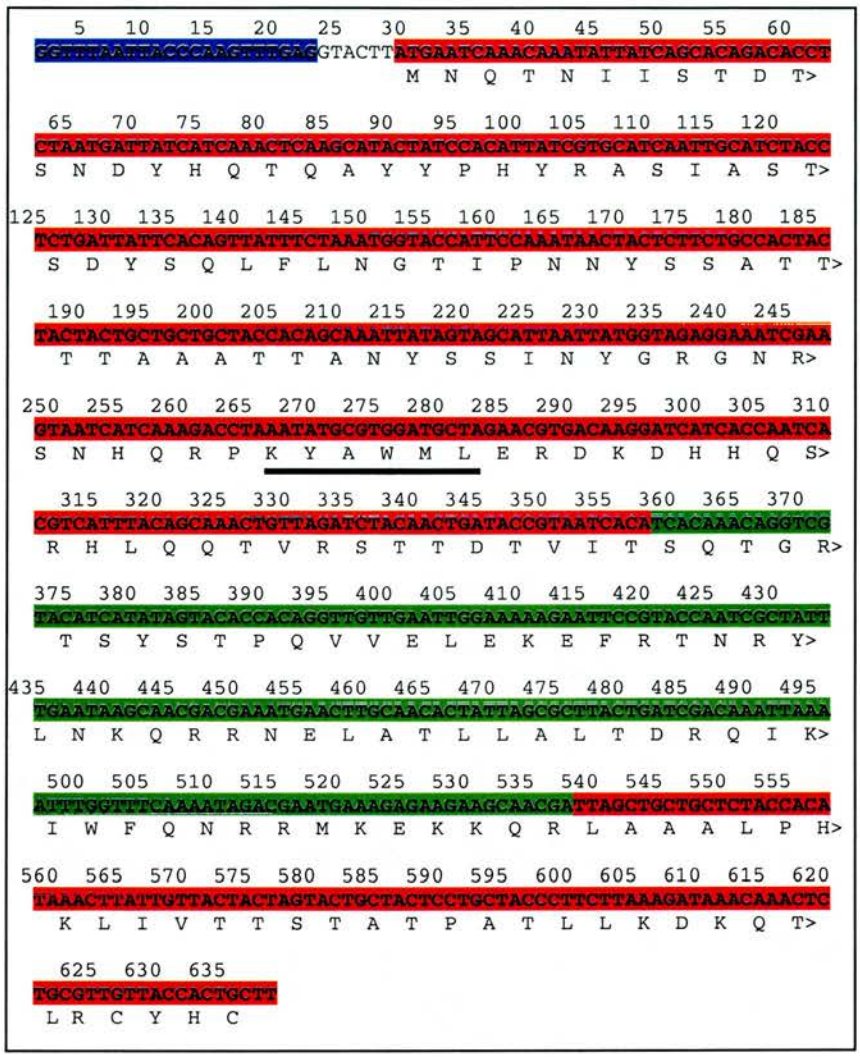


Fig 3.5 A partial *Bm-hox-3* cDNA clone

The *Bm-hox-3* partial clone is complete at the 5' end but remains truncated at the 3' end. The homeodomain region is highlighted in green with the rest of the coding region highlighted in red. A divergent hexapeptide is detectable in the sequence underlined by a black line. The SL1 leader sequence is highlighted in blue.

3.2.4. Demonstration of exon sharing by two different homeodomains in Brugia malayi

When 5' RACE was attempted for the genes *Bm-egl-5* and *Bm-ant-1* fragments of approximately the same size were amplified from first strand cDNA. Cloning and sequencing of the fragment generated using *Bm-egl-5* homeodomain specific reverse primers within the homeodomain and a 5' RACE primer revealed the fragment was the 5' end of this gene including the SL1 splice leader sequence before the beginning of the open reading frame. The 5' end of Bm-EGL-*5* has regions of homology to the *C. elegans* ortholog outside of the homeodomain (Fig 3.6). The 3' of the *Bm-egl-5* gene generated by 3' RACE is shorter than that of the *C. elegans* gene with little homology outside of the homeodomain (Fig 3.6). Subsequent amplification of *Bm-egl-5* fragments with reverse primers in the homeodomain and the SL1 splice leader primer indicated that the full length *Bm-egl-5* gene was SL1 trans-spliced. Furthermore, no internal splice sights were identified for this gene in contrast to that identified for *Bm-mab-5* gene.

Cloning and sequencing of the *Bm-ant-1* 5' RACE fragment gave a most surprising result. The 5' of *Bm-ant-1* was identical to that of the 5' end of *Bm-egl-5* at the nucleic acid level from the fifth amino acid N-terminal to the homeodomain to the start methionine (Fig 3.7). This fragment also had the SL1 splice leader sequence at its 5' end. This finding immediately suggested that these two Hox genes, belonging to different orthology groups according to the sequence of their homeodomains, shared the same 5' exon (Fig 3.7).

Further PCR based analysis on first strand cDNA using primers specific to the shared 5' region and homeodomain specific primers for each gene revealed that the 5' RACE results were not spurious (Fig 3.8). Both transcripts could be detected in adult female first strand cDNA. The homology of this 5' region to the 5' of *Ce-egl-5* suggests that the *Bm-ant-1* may be alternately spliced in place of the *Bm-egl-5* homeodomain. Further 5' RACE experiments for *Bm-ant-1* failed to identify any other 5' end for this gene, indicating that this gene has lost its own 5' end. Alternatively it is possible that the *egl-5* gene has adopted the shared 5' end originally belonging to the ancestral *Bm-ant-1* gene, which has subsequently been

lost from *C. elegans*. However, the putative divergent hexapeptide of the posterior group genes characterised by a tryptophan residue 6 or 7 residues from the start of the homeodomain is present in the *egl-5* genes (Burglin, 1994).

To begin investigating the genomic organisation of these genes PCR on *B. malayi* gDNA was performed. PCR designed to amplify the *Bm-egl-5* gene from gDNA produced a single product of approximately 1.2 kb. This product was cloned and sequenced and found to represent a genomic clone of the *Bm-egl-5* gene from the 5' terminus to the centre of the homeodomain (called Bm-egl5/1.2) with a single intron placed between the 5' exon shared by both genes and the *Bm-egl-5* homeodomain. This result confirms that the shared exon and the *egl-5* homeodomain are in cis in the genome. Similar attempts to amplify a similar genomic clone for the *Bm-ant-1* gene from gDNA failed. Similarly the *Bm-ant-1* homeodomain was not detected on the egl5/1.2 clone by PCR. This indicated the *Bm-ant-1* homeodomain exon did not lie between the shared 5' exon and the *Bm-egl-5* homeodomain exon. This was later confirmed by sequencing of the egl-5/1.2 clone including almost all the very A/T rich intron.

To see if the *Bm-ant-1* homeodomain was in cis with *Bm-egl-5* homeodomain PCR on gDNA between these two domains was also attempted using forward and reverse primers in all combinations from both genes. This approach again failed to produce any PCR products and was therefore uninformative as to the position of the *Bm-ant-1* homeodomain.

The egl5/1.2 clone was then used as a probe for hybridisation screening of a gridded *B. malayi* BAC library. The egl-5/1.2 clone was used as a template for generating a biotin labelled probe by PCR. Hybridisation to 4,608 gridded clones of average insert size of 50kb led to the identification of a single BAC clone. This clone was shown by PCR to be positive for the *Bm-egl-5* homeodomain, the shared 5' exon and the *Bm-ant-1* homeodomain. Further PCR analysis of this BAC clone found that the *Bm-ant-1* homeodomain lies downstream of the *Bm-egl-5* homeodomain, and is therefore located in cis to the shared exon as well.

Fig 3.6 Alignment of full length EGL-5 proteins from *B. malayi* and *C. elegans*

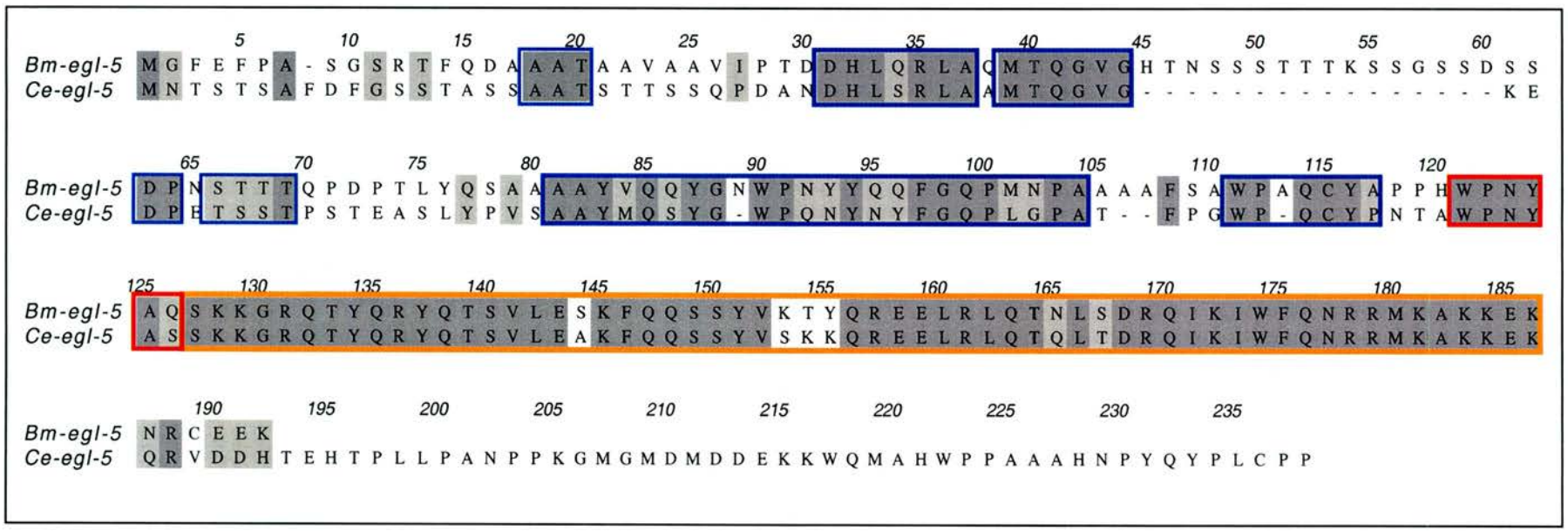


Fig 3.6 Alignment of full length EGL-5 proteins from *B. malayi* and *C. elegans*

An alignment of the full length EGL-5 proteins from *B. malayi* and *C. elegans*. The homeodomain region is boxed in orange and conserved flanking residues are boxed in red. Other regions of significant conservation are boxed in blue confirming that the 5' end of *Bm-egl-5* is indeed homologous to the 5' end of the *Ce-egl-5* gene. Grey residues indicate identities and light grey shading indicates similarities.

Fig 3.7 Alignment of Bm-EGL-5 and Bm-ANT-1 full length proteins

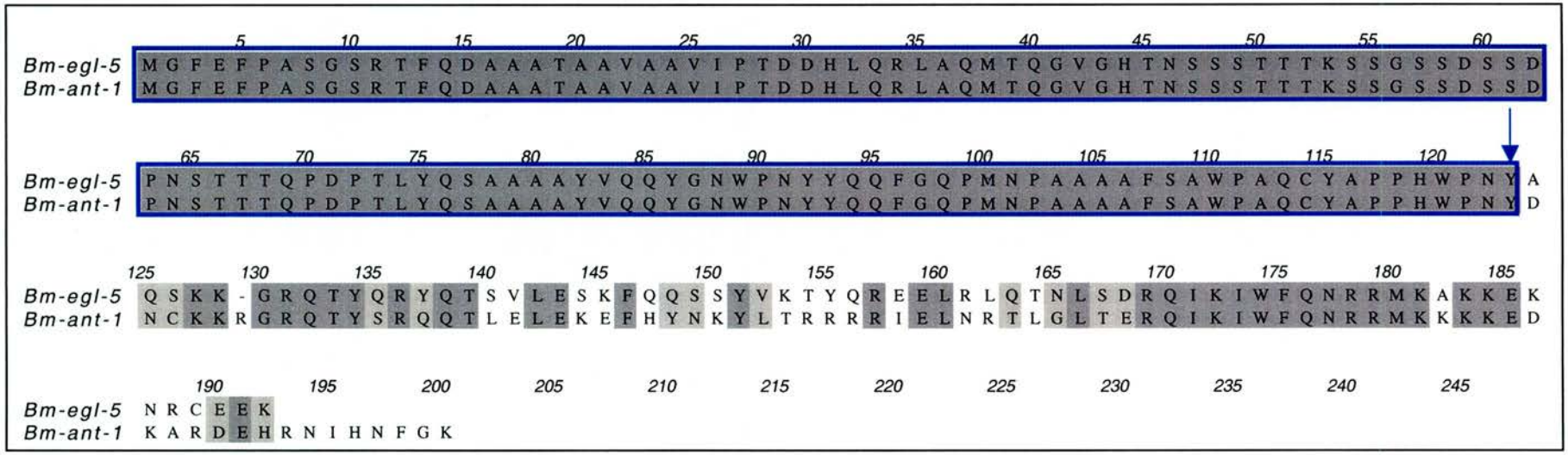


Fig 3.7 Alignment of Bm-EGL-5 and Bm-ANT-1 full length proteins

Alignment of the Bm-EGL-5 and Bm-ANT-1 proteins showing that they share the same 5' exon. The blue box highlights the translation of the shared 5' exon. The blue arrow indicates the position of the intron in the egl-5/1.2 clone and the point at which nucleic acid identity stops in the two cDNAs.

Fig 3.8 PCR of Bm-egl-5 and Bm-anl-1 using the same 5' primer and homeodomain specific primers for each transcript

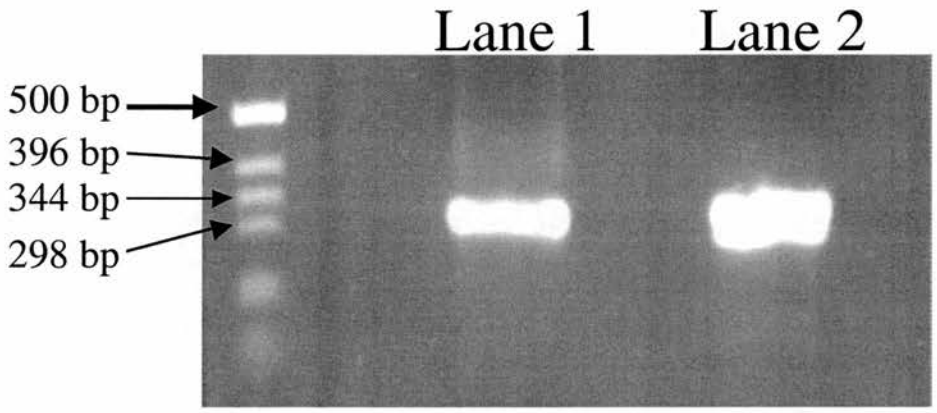


Fig 3.8 PCR of Bm-egl-5 and Bm-anl-1 using the same 5' primer and homeodomain specific primers for each transcript

PCR products from adult female cDNA showing that Bm-egl-5 (Lane 1) and Bm-anl-1 (Lane 2) transcripts share the same 5' exon. PCR reactions were performed with the same forward primer and primers specific to the homeodomain of each transcript. Sequencing of these products confirmed that they corresponded to the correct transcripts.

The BAC has been sized using Pulse Field Gel Electrophoresis and is approximately 26 kb. This clone has been submitted for sequencing to the Pathogens Sequencing Unit, Sanger Institute, Hinxton, Cambridge. One immediate point of interest is whether there is another coding exon representing the true 5' exon of *Bm-ant-1* that has not been detected by 5' RACE.

The space between the *Bm-ant-1* and *Bm-egl-5* homeodomains is approximately 2.4 kb analysed by PCR on DNA prepared from the BAC clone. In *C. elegans* the *egl-5* and *mab-5* genes are relatively closely linked relative to the other Hox genes along chromosome three of the model organism, about 28 kb apart, with a non-Hox gene in between. The BAC has provided the first evidence for the Hox linkage in a nematode distantly related to *C. elegans*. The presence of the *Bm-ant-1* homeodomain in close proximity to *Bm-egl-5* is tentative evidence to support the inclusion of *Bm-ant-1* in an orthology group physically closer to the posterior Hox genes than that represented by *Bm-mab-5*. The *Bm-mab-5* gene is not present on the BAC. By this argument we would expect the *mab-5* group to be an ortholog of the *fushi-tarazu* genes of Arthopods and *Bm-ant-1* and *As-ant-1* to represent an ortholog of the *Antennapaedia* group. In addition the *B. malayi* ortholog of the *Sr-ant-1*, suggested to be a *Lox2/Ubx* ortholog gene may have been lost from its position in the *B. malayi* cluster or lost altogether as it appears not to be between *Bm-egl-5* and *Bm-ant-1*. This possibility will only be confirmed when the intergenic region has been sequenced. If this is true then it represents a separate movement/loss event from those that have occurred in the lineage to *C. elegans* because the *S. ratti* and *B.malayi* lineages diverged before the *S. ratti* and *C. elegans* lineages.

The finding that these two homeodomains share the same 5' end has a number of implications for the functional evolution of these genes. This phenomenon seems to represent a novel type of molecular evolution for Hox genes, although there is on report of what may be a similar phenomenon for Hox genes on human chromosome 12 (Simeone et al., 1988). Although there have been no further reports on this phenomenon and the shared 5' region was non-coding (Simeone et al., 1988). The main implication is that these genes can interact with the same set of co-factors but bind different downstream targets as the DNA binding specificity of their homeodomains is likely to be different.

3.3. Preliminary analysis of Hox gene expression levels through the lifecycle of *Brugia malayi*

The lifecycle of the filarial parasite *B. malayi* can be maintained in the laboratory using mosquito and jird hosts. Microfilaria (Mf) are collected from the blood of sacrificed jirds and blood fed to mosquitoes. Within the mosquito host microfilaria develop through two larval stages in the mosquito flight muscles. After 12 days the Mf have matured to the infective third larval stage and are competent to re-infect a mammalian host and travel to the mosquito mouthparts. The L3 larvae are extracted from mosquitoes and injected into jird hosts where they develop into mature adults and subsequently release microfilaria into the host bloodstream (McVay et al., 1990).

In order to assess the expression levels of Hox genes through the *B. malayi* lifecycle (Fig 5.3 in chapter 5) first strand cDNA was prepared from pooled embryos, pre-blood feed microfilaria (Mf), mosquitoes from day 1 to day 12 post blood feed positive for *B. malayi* infection and adult females. Total RNA was prepared from each lifecycle stage, quantified using a spectrophotometer and used to make first strand cDNA. First strand cDNA was then used for PCR using primers to *B. malayi β-tubulin* (*Bm-tub-1*) and *Bm-ama-1* (the name of the gene corresponding to the large subunit of RNA polymerase II, (Bird and Riddle, 1989)). Amounts of cDNA to be used in subsequent PCRs from each lifecycle stage were normalised for transcript level of these two housekeeping genes that are relatively constant through the lifecycle (Gomez-Escobar et al., 2000). This was done by ascertaining the amount of first strand cDNA required to produce approximately the same amount of PCR product for each lifecycle stage for both these transcripts under standardised PCR conditions (see Materials and Methods section for details). This step was necessary as the majority of RNA from lifecycle stages prepared from mosquitoes is obviously from the insect. As the number of live parasites decreases during the 12 days of infection the ratio of mosquito RNA to parasite RNA increases in samples prepared from the end of this time-course.

PCR for each Hox gene transcript was then performed with specific primers amplifying fragments of approximately the same size using plasmid templates for

each gene at varying concentrations. PCR conditions were established that gave an easily visible difference between serial dilutions of plasmid template. These PCR conditions were then used to assess the levels of each Hox gene transcript at each developmental stage using the previously established normalised levels of first strand cDNA.

This method for assessing transcript levels gives a rough guide to transcript level through the lifecycle but is not as accurate as quantitative PCR techniques using radioactivity or fluorescence. However, the data described here provides a guide to more quantitative studies using these techniques and will act as a guide for choosing lifecycle stages for performing *in situ* analysis of the spatial expression pattern of *B. malayi* Hox genes. Here the preliminary assessment of the temporal expression patterns of six *B. malayi* Hox genes is presented. In addition PCR against a cDNA library prepared from the L2 stage of the parasite dissected out from the mosquito host was also used to assess the expression pattern of *B. malayi* Hox genes. Expression in this cDNA is not normalised in the same way as that for expression from first strand cDNA preparations. The expression pattern of *Bm-php-3* has not yet been assessed by this method.

Embryonic cDNA was prepared from embryos collected from the dissected gonads of adult female parasites. *B. malayi* adult females give birth to fully developed Mf (Rogers et al., 1976). Thus cDNA prepared from this stage will include transcripts from unfertilised oocytes, embryos at all stages of development and from pre-released Mf. The amount of pre-released Mf was limited by collection of material from only the medial and posterior region of the gonad where early embryogenesis takes place (Rogers et al., 1976). Nonetheless it is likely that this stage includes some material from pre-release Mf.

3.3.1. Expression profile of the Bm-ceh-13 transcript

The *Bm-ceh-13* transcript was detected in embryonic cDNA, pre-blood feed Mf cDNA and also at lower levels in cDNA from day 1 in the mosquito host. Transcript was also detected in adult female cDNA and is probably reflective of embryonic

expression (Fig 3.9). The transcript was not detected at any other stages during the development in the mosquito host. Consistent with this *Bm-ceh-13* was also absent form the L2 cDNA library (Fig 3.9). These data suggest a role for *Bm-ceh-13* in embryogenesis and early development but not in later stages of parasite development. In *C. elegans ceh-13* plays a role in embryogenesis specifying many different cell fates and its absence results in defects in gross morphology and defects in normal cell lineage (Brunschwig et al., 1999; Wittmann et al., 1997b).

3.3.2. Expression profile of the Bm-hox-3 transcript

The *Bm-hox-3* transcript was detected in embryos, pre-blood feed Mf and in day 1 post blood feed mosquitoes at high levels. Transcript level then appears to drop from days 2-4 but is still detectable rising again in days 5 and 6 (Fig 3.9). Transcript is also detectable at days 7 and 8 at low levels and then is absence for the rest of the time course. As parasites in mosquitoes do not develop in exact synchrony the expression detected either side of days 5 and 6, in between the two moults may be due to asynchronous development of some parasites rather than reflect lower expression levels in all parasites. Expression is also detected in the L2 cDNA library consistent with higher expression levels on days 5 and 6, and in adult female cDNA consistent with the embryonic expression. As this gene is absent from *C. elegans* no preliminary comparison with a closely related organism is possible.

3.3.3. Expression profile of the Bm-lin-39 gene

The *Bm-lin-39* expression profile is very similar to that described for the *Bm-hox*-3 transcript. Expression levels are high in embryonic cDNA, pre-blood feed Mf and for days 5, 6 and 7 of the mosquito time course and then disappears. Expression is detected but lower for days 2-4 (Fig 3.9). Consistent with embryonic expression the transcript is also detectable in adult female cDNA. One inconsistency however is that the transcript could not be detected in the L2 cDNA library under the standard PCR conditions deduced for the *Bm-lin-39* gene. As expression is detected during the time

course at an equivalent stage this is surprising. The transcript could be detected in the library at very low levels if 35 cycles of PCR were used as opposed to the 29 cycles for the standard PCR (see Materials and Methods section.).

The *lin-39* gene in *C. elegans* plays a major post-embryonic role in specifying and then controlling vulval development. In this context it has expression peaks at the end of embryogenesis and at the L1 stage and then again at the L3 stage when vulval development is induced (Maloof and Kenyon, 1998). If the *Bm-lin-39* gene plays a similar role in this nematode it would suggest that vulval induction occurs not at he L3 stage but at the L2 stage of development. However only analysis of spatial expression and perhaps subsequent functional analysis will confirm this hypothesis.

3.3.4. The Bm-mab-5 transcript is on, switches off and then on again

The *Bm-mab-5* has a very distinct expression profile. It is expressed at high levels in embryonic cDNA, pre- blood feed Mf and in day 1 of the mosquito time course (Fig 3.9). It then remains of until switching on again at high levels for days 5 and 6, with lower expression on days 4 and 7 likely to represent asynchronous parasite development. Expression is also detectable in adult female cDNA and in the L2 cDNA library consistent with the expression profile.

The *mab-5* gene in *C. elegans* has many different functions in specifying cell fate so it is difficult to make any comparison at this stage. At this stage the expression profile of only the longer *Bm-mab-5* transcript has been assessed. It will be informative to see if the shorter transcript has a different expression profile. RT-PCR analysis encompassing only the shorter transcript may reveal it is expressed at stages when the longer transcript is not. However detecting when the longer transcript is expressed and the shorter transcript is not will require an alternative technique such as Northern blot analysis of expression.

3.3.5. Expression profiles of the Bm-ant-1 gene and the Bm-egl-5 gene are overlapping but distinct

As described above these two Hox genes appear to share the same 5' exon suggesting a novel form of molecular organisation for Hox genes. An assessment of the expression profile of these two genes through the *B. malayi* lifecycle indicates that they are distinct but overlapping (Fig 3.9).

The *Bm-egl-5* gene is expressed in a pattern very similar to that of *Bm-mab-5*. It is present in embryonic cDNA, pre-blood feed Mf cDNA and day 1 post-blood feed cDNA identical to the pattern seen for *Bm-mab-5* (Fig 3.9). The transcript is not detected again until day 5, remains detectable through days 6 and 7 at lower levels and then expression is not detectable. This expression pattern is similar to that of *Bm-mab-5* but appears to switch on slightly later as it is not detected on day 4. Consistent with the results for the mosquito time course the transcript is also detected in the L2 cDNA library as well as adult female cDNA.

In contrast the *Bm-ant-1* transcript incorporating the *Bm-ant-1* homeodomain and the same 5' exon as the *Bm-egl-5* genes is restricted to earlier stages of development. This transcript is only detected embryonic cDNA, pre-blood fed Mf cDNA and adult female cDNA. The transcript is not detected at any other point in development in the lifecycle stages investigated here. This suggests that this transcript may have a role early in development but not at other stages.

Thus the expression patterns of these two transcripts are regulated very differently even though they share the same 5' exon.

Fig 3.9 RT-PCR expression analysis of Hox genes through the *B. malayi* lifecycle

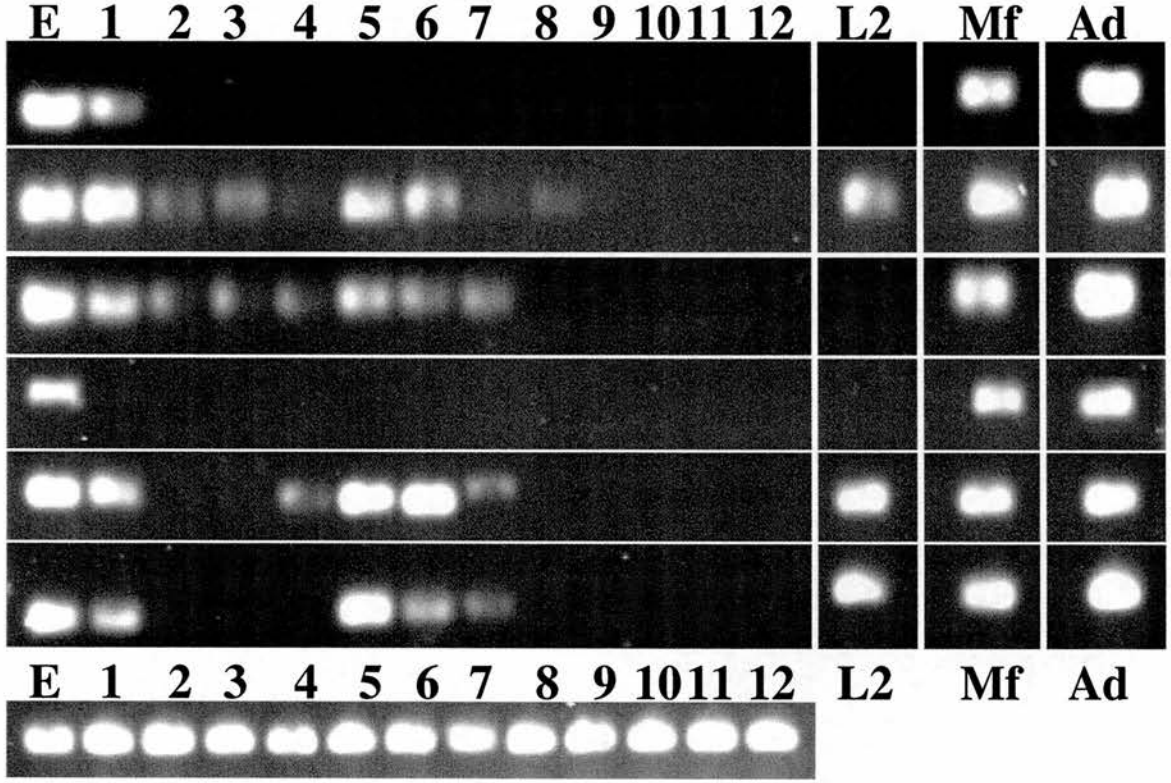


Fig 3.9 RT-PCR expression analysis of Hox genes through the *B. malayi* lifecycle

The transcript levels of *B. malayi* Hox genes were assessed through the lifecycle of the parasite. RT-PCR was performed on cDNA prepared from embryonic, pre-blood feed Mf, mosquito life cycle stages and adult females. The amount of cDNA from each lifecycle stage was normalised across all stages by using the transcript levels of two the housekeeping genes *Bm-tub-1* and *Bm-ama-1*. Standardised PCR conditions were established empirically for each Hox gene against serial dilutions of cloned inserts for each gene. In addition transcript level in an L2 stage cDNA library were also assessed (see Fig 5.3 for *B. malayi* lifecycle).

3.4. Cloning of regions outside of the homeodomain from other nematodes

Attempts have been made to clone regions outside of the homeodomain from other nematodes using 5' and 3' RACE. However. For the most part due to limits in the amount of material available this has not been very successful. Material from *S. ratti* and *M. javanica* in particular was not ideal and the RNA produced from these samples was of poorer quality than hoped. Future attempts at cloning the 5' and 3' ends from the other species will be made when more material is available. This will be particularly important for genes where flanking residues will help to prove of disprove the orthology currently assigned to them, such as the *Sr-ant-1* gene Nonetheless, for some genes it was possible to clone some extra sequence beyond that identified by the original degenerate PCR approach if not full length 5' or 3' ends. The *As-ant-1* one sequence corresponds to an unpublished sequence referred to by Schaller et al 1990. This homeodomain was cloned from *A. suum* and used to screen *C. elegans* cosmid libraries used in the *C. elegans* genome project. The full homeodomain sequence was made available by Dr Thomas Burglin and has been used in the phylogenetic analysis in chapter 4.

Small 5' extensions were cloned for the *Sr-ceh-13*, *As-ceh-13*, *Ts-hox-3*, *As-hox-3* and *Sr-php-3*. More extensive 5' extensions were cloned for *Ts-ant-1* and *Ts-ant-3* that encompassed the full N-terminus of the homeodomain but did not stretch as far as the hexapeptide motif. These extra sequences have already been included in the alignments in chapter 2. Sequences outside of the homeodomain for *Ts-ant-1* and *Ts-ant-3* are not informative with respect to orthology assignment.

The C-termini of *Ts-hox-3*, *Ts-lin-39*, *Ts-ant-1*, *Ts-ant-3* and *Ts-php-3* were identified by 3' RACE. The 3' ends of these sequences beyond the homeodomain do not have any informative regions of homology to either genes from other phyla or orthologous nematode genes.

All the nucleic acid sequences generated are presented in Appendix 1.

3.5. Towards studying the role of Hox genes in Nematodes only distantly related to *C. elegans*

The preliminary work described in this chapter will form the basis for further analysis of Hox genes from *B. malayi* It should be possible to address issues of expression pattern, genomic organisation and possibly function for the *B. malayi* Hox genes. In addition it should be possible to employ *C. elegans* as a heterologous transgenic system to investigate the evolution of both regulatory and coding regions.

Two Hox genes have been found in *B. malayi* that are not present in the genome of *C. elegans*. Investigating the expression pattern of these genes in comparison to those genes present in *C. elegans* should provide an insight into the evolution of the Nematode body plan. It should also be possible to take putative promoter regions from these genes and see if they are functional in *C. elegans*. It would be intersting to see if these regions are able to limit the expression of transgenes to regions of the *C. elegans* anteroposterior axis corresponding to their assumed positions according to spatial colinearity. It will also be interesting to see in expression of the missing Hox genes is able to in any way change cell fates during *C. elegans* development or partially substitute for the function of any of the remaining Hox genes.

The finding that two different homeodomains share the same 5' exon in *B. malayi* appears to be a novel mechanism of Hox gene expression. Along with gene loss, an increased evolutionary rate (see chapter 4) and loss of colinearity it represents another way in which the nematode Hox genes are different from those of other phyla. The main functional implication of this observation is that these two homeodomains can bind different downstream targets while interacting with the same partners. The preliminary analysis of expression levels through the lifecycle for both *Bm-egl-5* and *Bm-ant-1* suggests that these genes are expressed at different times during parasite development. *Bm-ant-1* is expressed only in early stages while *Bm-egl-5* is also expressed at later stages (see Fig 3.8). We would also expect the spatial expression pattern of these genes to be different form each other as one is central class gene (see chapter 4) and the other a posterior class gene. The role of *Ce-*

egl-5 is well understood in *C. elegans* and it should be possible to compare the function/expression pattern of *Bm-egl-5*.

It remains to be seen whether or not *Bm-ant-1* has its own 5' end or not but genomic sequencing of the BAC clone containing both homeodomains should help to clarify this. Constructing a genomic clone containing the *Bm-egl-5* gene and *Bm-ant-1* gene and flanking regions to use a transgene construct into *C. elegans* may indicate whether the model nematode resolves the transcripts in the same way. The partial cloning of *As-egl-5* and *As-ant-1* will allow the testing of the conservation (if any) of the organisation observed in *B. malayi*. Cloning the 5' ends of these from *A. suum* genes as well as investigating their genomic organisation may indicate whether the sharing of 5' ends in *B. malayi* has occurred very recently in evolution or is more widespread in clade III of the Nematoda.

Further studies of *B. malayi* Hox genes and those of other nematodes, linked to a better understanding of developmental mode in other nematodes (eg cell lineage analysis of nematodes only distantly related to *C. elegans*), will allow the testing of specific hypotheses of Hox gene evolution in the Nematoda. If rapid evolution and gene loss are concomitant with a movement towards a cell lineage dependent mode of development this should be observable through the phylum. The study of Hox gene function might then be expected to show stages of intermediate redundancy and reemployment that are already observed in *C. elegans*, as well as gene loss. If *C. elegans* is representative of the most extreme extant state of this process we might expect other nematodes to have retained more Hox genes, with their functions being closer to the ancestral function of specifying positional identity not cell lineage identity. One prediction of this proposal is that these nematodes would not have an invariant cell lineage during embryogenesis.

3.6. Acknowledgements

I would like to thank Ms Jennifer Daub for her training in the screening of gridded BAC genomic libraries. I am also grateful to Ms Claire Whitton for training me in the use of the Pulse Field Gel apparatus and sizing the BAC clone. Finally thanks to

the DNA Sequencing Lab at NEB for providing biotin labelled primers and to Professor Barton Slatko and Ms Betty Slatko for their hospitality during my stay in New England.

CHAPTER 4: A PHYLOGENETIC ANALYSIS OF HOX GENES FROM NEMATODES.

4.1. Phylogenetic analysis of Hox genes

Despite their central conserved role in the evolution of animal body plans at the micro-evolutionary and macro-evolutionary scale Hox gene sequences are not ideal tools for phylogenetic analysis. The short 60 amino acid DNA binding region is highly conserved and can be aligned unambiguously allowing phylogenetic analysis that places Hox genes in deuterostomes and protostomes into orthologous groups (de Rosa et al., 1999), shared across many phyla, and identifies some that may have arisen by independent gene duplications within specific lineages (see chapter 1). Reconstructing the pattern of Hox gene duplication and loss and sequence evolution using the short conserved DNA binding region can result in ambiguities that can only be resolved by employing other methods, if at all. This is compounded by the fact that many identified Hox homeodomains are only partial sequences that have not been extended beyond fragments cloned in degenerate PCR screens (Long and Byrne, 2001). Nonetheless these studies are important in forming the basis for our picture of body plan evolution at the molecular level, and in the absence of functional and linkage studies, form the only currently Hox gene data for some taxa.

The use of characteristic peptide motifs outside of the homeodomain and characteristic residues within the homeodomain have been useful, providing an alternative test for assigning genes to orthology groups in the absence of significant overall phylogenetic signal (Balavoine, 1997; de Rosa et al., 1999; Kobayashi et al., 1999; Telford, 2000c). Phylogenetic analyses incorporating residues flanking the homeodomain attempt to increase the amount of phylogenetic signal, but should only be used when residues can be aligned unambiguously without concern for the absence or presence of insertions and deletions (Ferrier et al., 2000). However, this is only possible when full-length homeodomains and flanking sequence is available, and for many genes in many taxa this information is still missing. Another phylogenetic technique that has been used to improve Hox gene phylogenetics is to use a combined methods approach incorporating a combination of distance,

parsimony and likelihood analyses to arrive at a consensus phylogeny (Kourakis and Martindale, 2000; Takahashi et al., 2001; Telford, 2000a). Using a combination of methods allows greater confidence to be attributed to any shared tree topologies independently generated by different methods that are or are not supported statistically (Kim 1993). Combined methods analysis allows the relationship between different Hox genes to be estimated with greater confidence and allows a more confident reconstruction of the evolution of Hox genes and ParaHox genes (Kourakis and Martindale, 2000; Long and Byrne, 2001).

Despite these refined methods the minimal phylogenetic signal present in Hox gene sequences means that sequence evolution can obscure orthology relationships across phyla and make the reconstruction of duplication events difficult to trace (de Rosa et al., 1999). The issue of true orthology relationships is especially important for the central Hox cluster genes of deuterostomes (*Hox6-8*) and protostomes (*ftz/Lox5*, *Antp*, *Ubx/Lox2* and *abd-A/Lox4*). The central protostome genes can be grouped into orthology arrangements (see chapter 1) due to the presence of characteristic residues within and flanking the homeodomain (de Rosa et al., 1999). However, while these Hox signatures are strongly represented within the some members of the Lophotrochozoa and Ecdysozoa, central class genes from other members of each group cannot be assigned to an orthology group with confidence (de Rosa et al., 1999). In these cases sequence evolution has obscured the confident designation of orthology (for example the central class genes of *Priapulus caudatus* apart from the *Ubx*). These genes may also represent lineage specific Hox gene duplication events followed by enough sequence evolution to obscures paralogy relationships. The orthology assignment problem is compounded by the fact that PCR screens are unlikely to identify all the Hox genes in a species. Complete Hox sets are only reliably found following the assembly of a genomic walk of the Hox cluster and/or significant genome sequencing (Ruvkun and Hobert, 1998). Perhaps the only practical way to tackle this problem is to sample Hox genes from a wider spectrum of extant phyla within the context of an accurate phylogenetic framework. The inclusion of more taxa may allow confident assignment of apparently divergent genes to orthology groups and by the same analysis allow likely lineage specific duplications (and losses) to also be identified. All the phylogenetic analysis in this

chapter is based on the alignment presented in NEXUS format in Appendix 3 (Fig 1) of this thesis and sub-alignments created from this major alignment. Phylogenetic methods used here are described in the Materials and Methods section.

4.2. Assignment of nematode Hox genes to paralogous groups by phylogenetic analysis

For some of the nematode Hox genes assignment to orthologous groups in relation to other phyla is trivial. However for others, particularly the central group genes assignment to orthologous groups is complicated by the sequence evolution that has occurred in the nematode lineage. The nematode *ceh-13*, *hox-3* and *lin-39* groups are likely to belong to the *Hox1/labial*, *Hox3* and *Hox4/Dfd* orthologous groups of other phyla respectively (Burglin 1994). The *php-3* nematode group is certainly an ortholog of the *Abd-B* gene found in other Ecdysozoa (de Rosa et al., 1999). This leaves uncertainty over the assignment of the central nematode group genes that have been identified. Grouping these genes within the Nematoda (see chapter 2) suggests that they represent at least three paralogous groups within the nematode lineage and at least two when compared to other phyla. Here phylogenetic analysis is used in an attempt to clarify this and also investigate the evolution of the other nematode paralogy groups in comparison to those of other phyla.

Recent studies of Hox gene phylogeny have included residues flanking the homeodomain when they can be aligned unambiguously (de Rosa et al., 1999; Telford, 2000a). The current nematode dataset does not contain any of the characteristic peptide motifs found in the flanking regions of Hox genes from other Bilaterians, so phylogenetic analysis has been performed with the 60 amino acid homeodomain or residues therein.

4.2.1. Assignment of genes to nematode paralogous groups

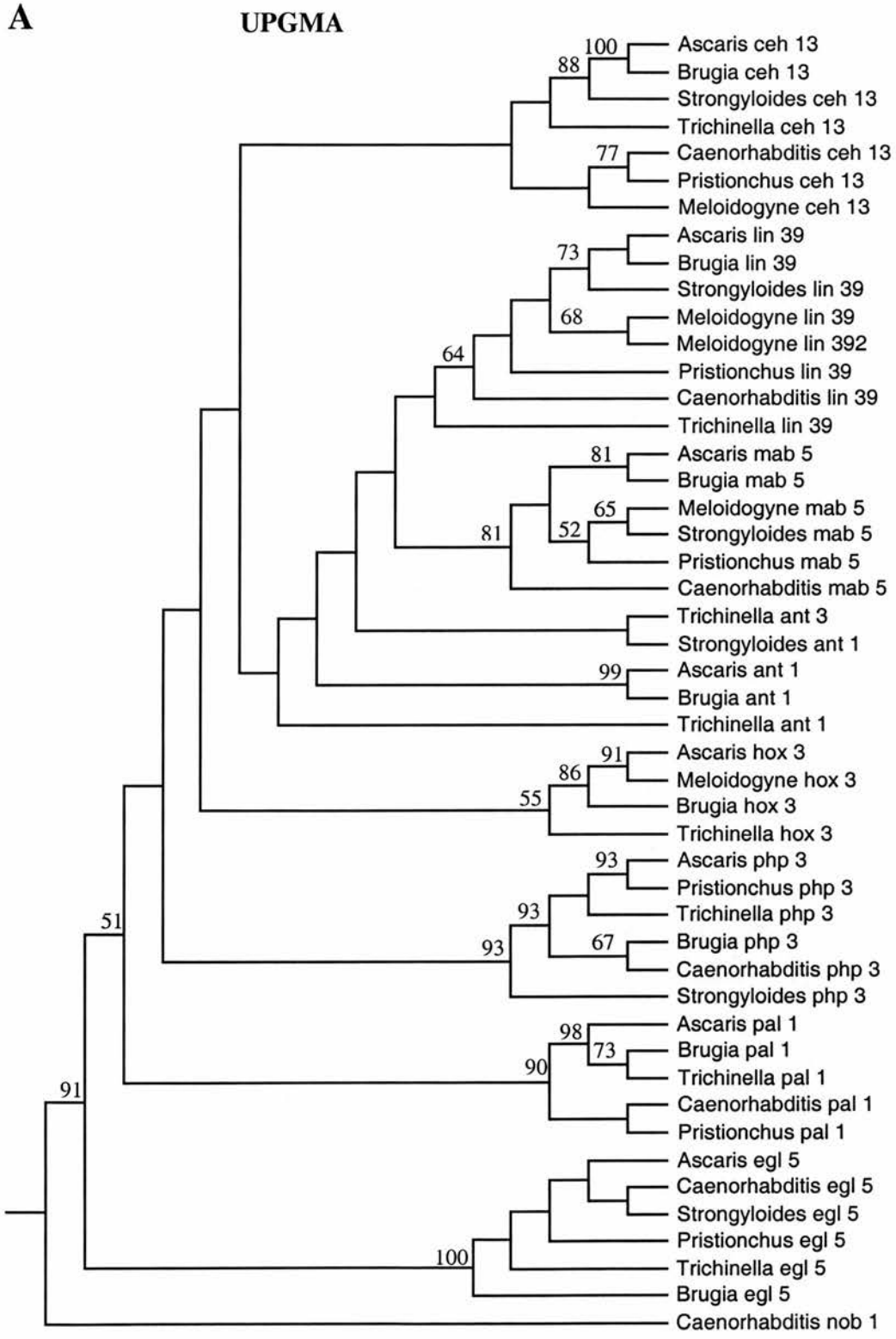
An initial phylogenetic analysis of nematode Hox genes was performed without including genes from other phyla. A combined analysis approach has been taken

using both distance and parsimony methods. All nematode Hox and ParaHox genes (represented by only the caudal family) have been included except for the second caudal gene, *Bm-cad-2,* discovered in *B. malayi* and the very divergent short homeodomain sequence of *Ts-ant-5*. This leaves an alignment of 48 nematode genes the relationships of which have been estimated by UPGMA (Unweighted Pair Group Method with Arithmetic means), maximum parsimony (MP), and neighbour-joining (NJ) methods (Fig 4.1 A,B,C, and D). It was hoped that this analysis along with the identification of characteristic residues provided in chapter 2 would provide a guide to the number of paralogous groups discovered in nematodes so far. In addition it could identify orthologous groups that may have arisen within the nematode lineage as oppose to being orthologous to Hox gene groups of other phyla. For example is it possible that the *Bm-ant-1* and *As-ant-1* genes are recent rather than ancient paralogs of the *mab-5* group? Ascertaining the relationship of genes within the Nematoda should identify which genes need to be analysed with those from other phyla to see if they are nematode specific paralogy groups or are directly orthologous to the paralogy groups of other phyla. Nematode Hox genes have been analysed along with the nematode *pal-1* caudal-like gene family (as an outgroup in NJ and MP analysis).

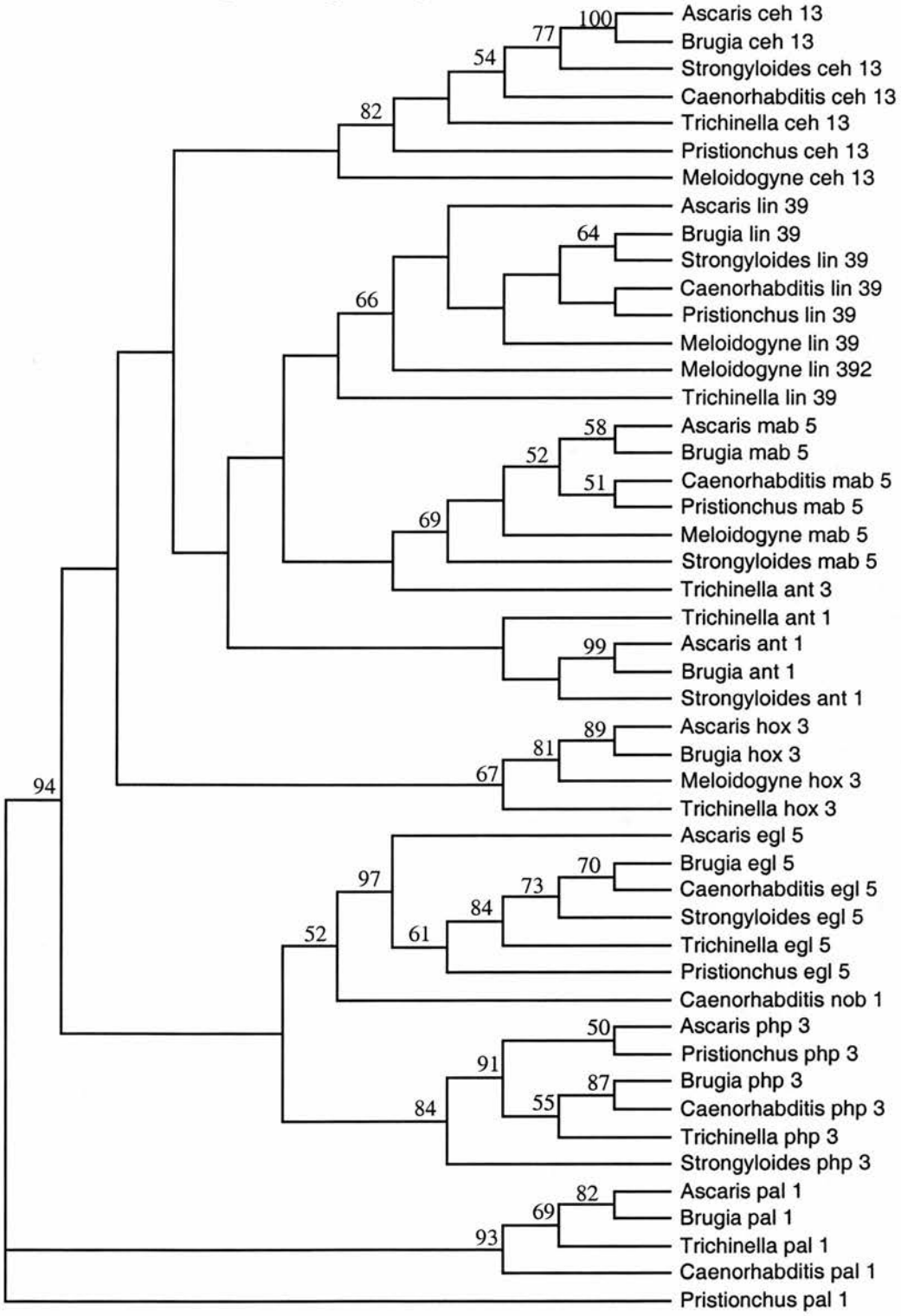
A combined analysis approach has clearly identified six Hox nematode paralogy groups (NPGs) and possibly up to two more central Hox gene groups making a possible total of eight (Fig 4.1 D). The six clearly identified NPGs are the *ceh-13* group, the *hox-3* group, the *lin-39* group, the *mab-5* group, the *egl-5* group and the *php-3* group. The genes assigned to these groups by characteristic residue analysis are recovered as being monophyletic with respect to the other nematode Hox genes by all three approaches. In some cases bootstrap support (>50%) is not observed although monophyletic groups are consistently recovered by all methods. The two exceptions to this observation are *Ts-lin-39* and *Mj-ceh-13*, which are not grouped with the other genes in their groups by parsimony analysis but are by Neighbour joining and UPGMA. The reasons for this appear to be because Mj-ceh-13 is a particularly fast evolving sequence (see below) and *Ts-lin-39* lacks characteristic residues specific to the *lin-39* group, and is closer to the *Dfd* genes of other Bilateria (see below).

Predictably, the assignment of the other central Hox cluster genes outside of the *lin-39* and *mab-5* groups remains unclear. *Bm-ant-1* and *As-ant*-1 are members of a different NPG to the *mab-5* genes as both these species also have cognate *mab-5* genes. These genes are strongly grouped in all analyses with strong bootstrap support and thus represent a seventh NPG. *Ts-ant-1* groups with these genes by NJ analysis and by parsimony analysis (with bootstrap support) and may be a member of this NPG. This grouping is not recovered by UPGMA analysis, however the assumption of equal evolutionary rates implicit in this method make it the least suitable for analysing nematode Hox genes which are evolving rapidly and at unequal rates (see below). The two remaining central genes, *Sr-ant-1* and *Ts-ant-3*, do not group with any of the other seven NPGs or with each other either by bootstrap support in any one method or by looking for any consensus topology between the three. They are closest to the *mab-5* NPG and the *Bm/As-ant-1* NPGs indicating that they are indeed central Hox genes. Between them they are likely to represent at least one more NPG and possibly two. Analysis of the central nematode Hox genes with those from other phyla could help clarify this position. Phylogenetic analysis to formalise the orthology assignments of each NPG to paralogous groups across the Bilateria is described below.

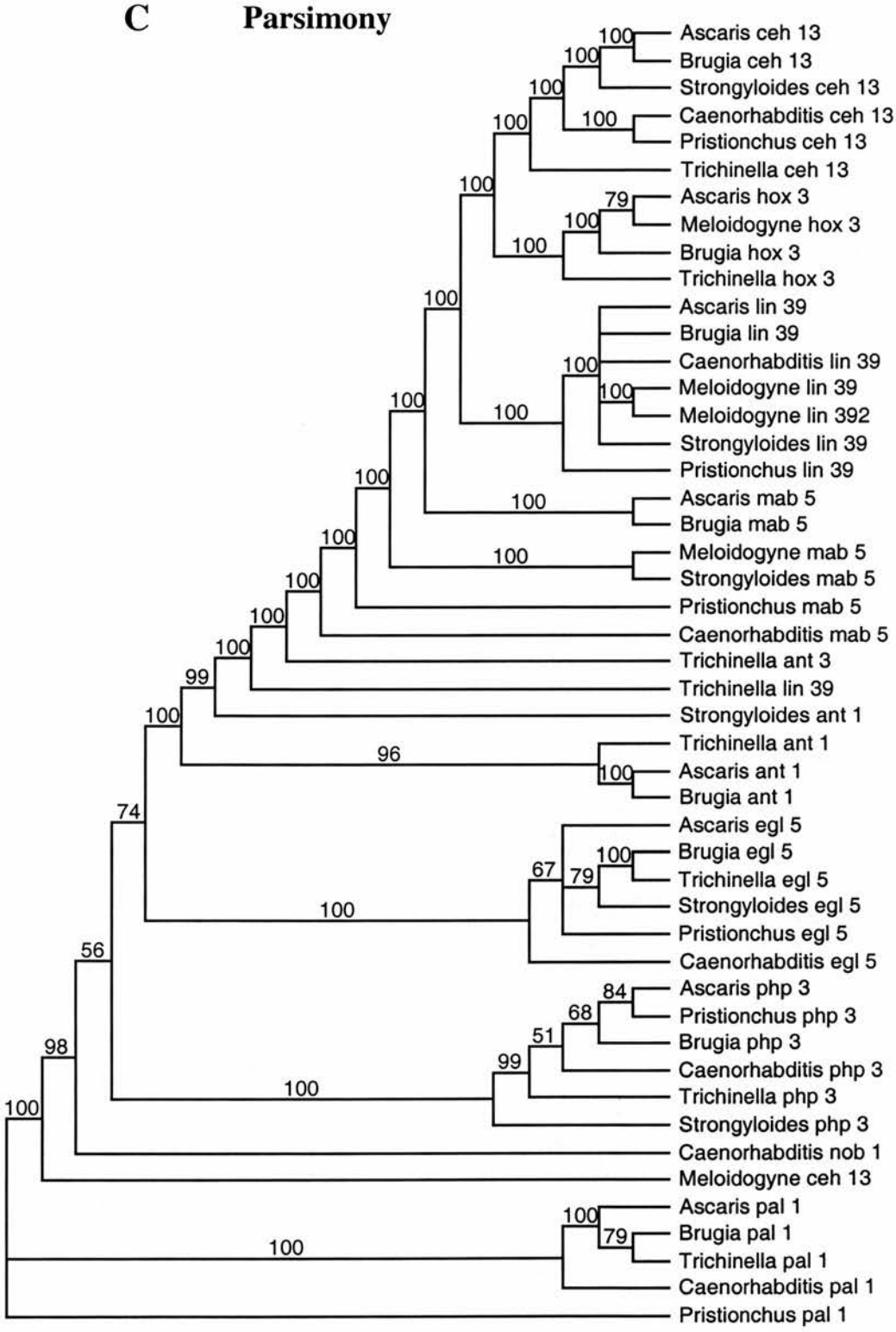
Figure 4.1. Phylogentic analysis of Hox genes within the Nematoda



B **Neighbour-joining**



C **Parsimony**


Ascaris ceh 13
Brugia ceh 13
Strongyloides ceh 13
Caenorhabditis ceh 13
Pristionchus ceh 13
Trichinella ceh 13
Ascaris hox 3
Meloidogyne hox 3
Brugia hox 3
Trichinella hox 3
Ascaris lin 39
Brugia lin 39
Caenorhabditis lin 39
Meloidogyne lin 39
Meloidogyne lin 392
Strongyloides lin 39
Pristionchus lin 39
Ascaris mab 5
Brugia mab 5
Meloidogyne mab 5
Strongyloides mab 5
Pristionchus mab 5
Caenorhabditis mab 5
Trichinella ant 3
Trichinella lin 39
Strongyloides ant 1
Trichinella ant 1
Ascaris ant 1
Brugia ant 1
Ascaris egl 5
Brugia egl 5
Trichinella egl 5
Strongyloides egl 5
Pristionchus egl 5
Caenorhabditis egl 5
Ascaris php 3
Pristionchus php 3
Brugia php 3
Caenorhabditis php 3
Trichinella php 3
Strongyloides php 3
Caenorhabditis nob 1
Meloidogyne ceh 13
Ascaris pal 1
Brugia pal 1
Trichinella pal 1
Caenorhabditis pal 1
Pristionchus pal 1

D Consensus of major groupings for UPGMA/NJ/MP trees

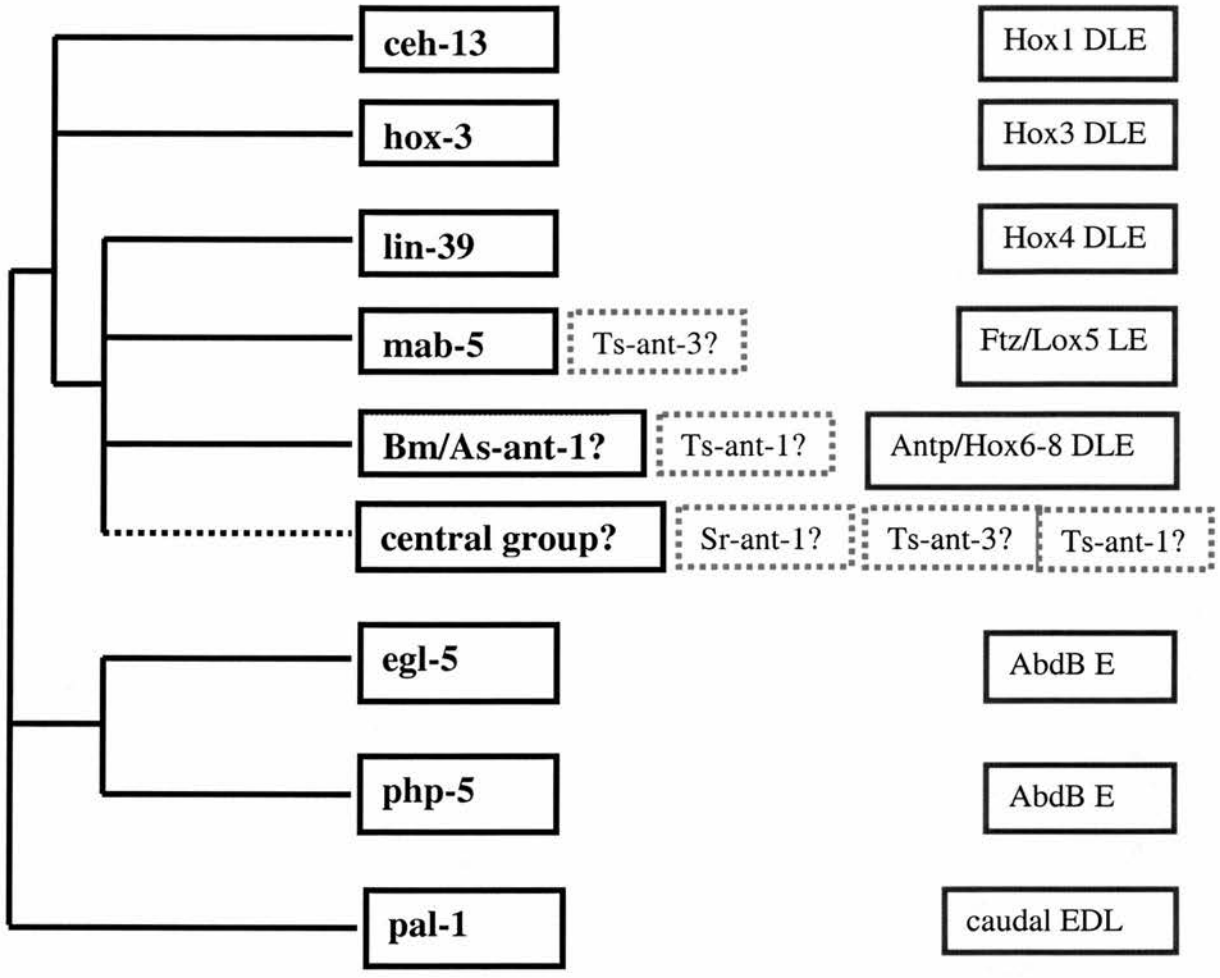


Figure 4.1. Phylogentic analysis of Hox genes within the Nematoda

Phylogenetic analysis of alligned nematode Hox protein sequences was peformed using PAUP 4.08b4. The 60 amino acid homeodomains or parts thereof were subjected to **A** UPGMA and **B** Neighbour joining distance (NJ) and to **C** Maximum Parsimony (MP) analysis. Where terminal residues were unknown they were treated as missing data. Values above nodes for trees depicted in A and B are bootstrap support over 1,000 replicates. Values above nodes in the tree depicted in C are the percentage of 10,000 most parsimonious trees that contained that node. **D** A diagramatic consensus tree was constructed from the three diferent analyses highlighting nematode paralogous groups (NPGs). Some central nematode genes did not clearly resolve into a NPG and these are named in red dashed boxes and placed next to putative NPGs in (black boxes) to which they may belong. Next to each NPG the abbreviated names of paralogous groups of major the Bilaterian linages (Deuterostome D, Lophotrochozoa L and Ecdysozoa E) to which they are proposed to be homologous are presented in green boxes. The presence of question marks indicates uncertainty in assignments. The *mab-5* group is proposed as a *ftz/Lox5*

ortholog (after (Telford, 2000a)). Nematode species and their relationships are described in chapter 2.

A full allignment for this analysis and full species names is presented in Appendix 3 (Fig 1 and Table 1).

4.3. Phylogenetic analysis of the *ceh-13*, *hox-3* and *lin-39* group genes of nematodes

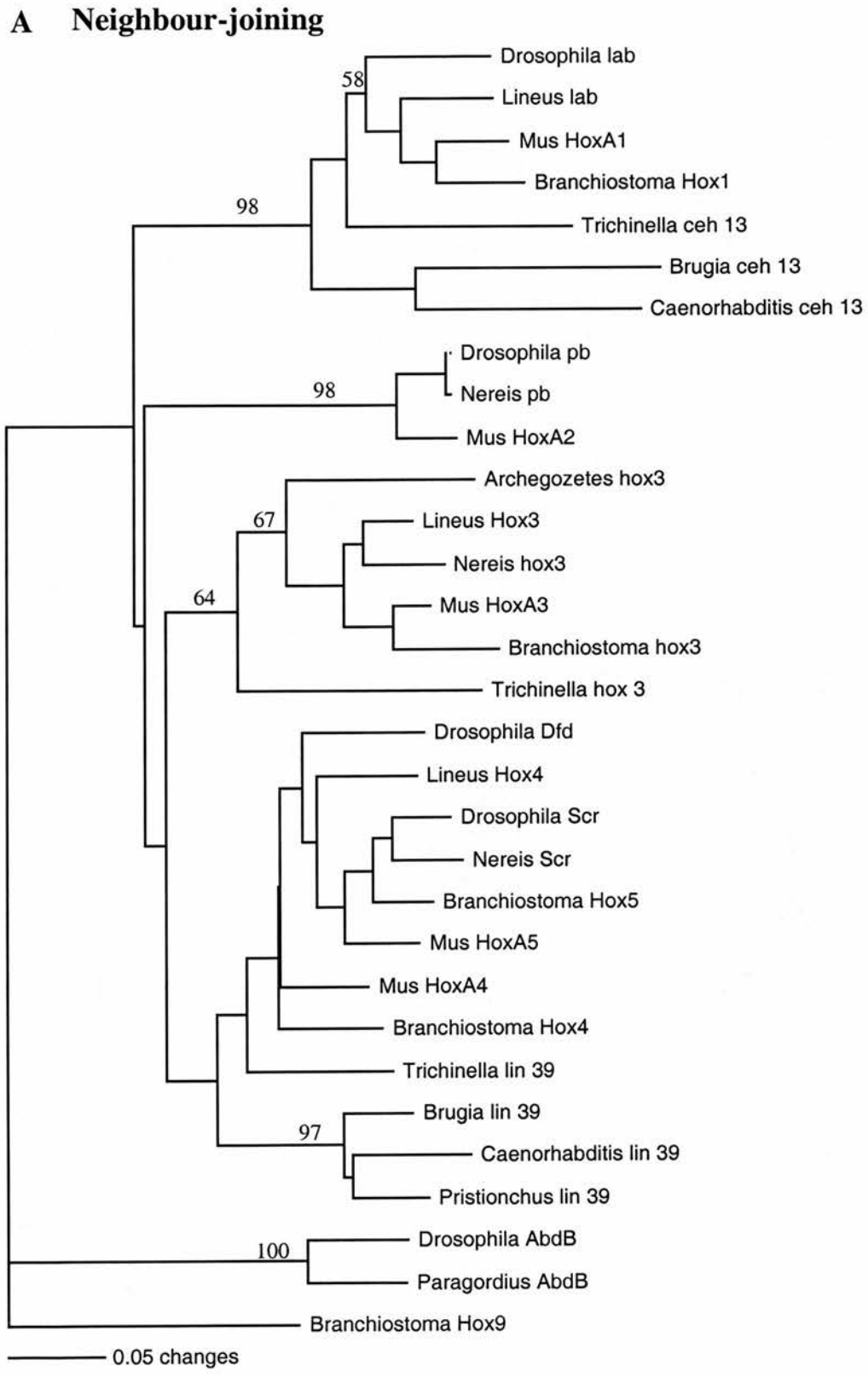
To confirm that the *ceh-13*, *hox-3*, and *lin-39* NPGs are orthologous to the *labial/Hox1*, *Hox3* and *Deformed/Hox4* groups respectively of other phyla phylogenetic analysis was performed with members of these groups from other bilaterians. Both NJ and MP analysis recovered monophyletic groups including the members of the NPGs and those of other phyla. For the *hox-3* group only the *T. spiralis* gene was included because the other members of this gene family in the Nematoda are even more divergent. For the other groups only nematode genes where the full homeodomain was known were included except for *Ts-ceh-13* and *Ts-lin-39* which were included because of the basal position of this species within the Nematoda.

By NJ analysis the *lin-39* genes were not monophyletic as the less divergent *Ts-lin-39* gene clustered with the main group of *Hox4/Hox5* genes. In this analysis the *Hox5* genes cluster within the *Hox4* group (Fig 4.2 A). Also by NJ analysis the *Ts-hox-3* gene is included in a group with the other *Hox3* genes included in the analysis. Parsimony analysis also includes the *Ts-hox-3* gene in this group with bootstrap support. All nematode *ceh-13* genes are included with the *Hox1* genes from other bilateria with bootstrap support in both NJ and MP analyses (Fig 4.2 A and 4.2 B). Both the *lin-39* and *ceh-13* nematode genes form an outgroup with bootstrap support from their orthologs in other phyla probably as a result of their divergent homeodomain sequences and residues characteristic of the Nematoda. The *Ts-lin-39* gene however does not group with the other *lin-39* genes by NJ or MP analyis, and as suggested in chapter 2 is more like the *Hox4* genes from other phyla. Most of the divergence in the *lin-39* sequences may therefore have evolved after the divergence of clade 1 of the Nematoda from the other lineages.

To ensure that the *Ts-hox-3* gene was a true member of the *Hox3* parology group a MP analysis including the *Branchiostoma floridae* (*Bfl*) ParaHox gene *Xlox* was performed (Fig 4.3). Analysis of this kind is advisable to check *Hox3* assignment in the absence of linkage data as ultimate support. The *Ts-hox-3* gene

grouped with other *Hox3* genes with bootstrap support with the *Bfl-Xlox* as an out-group, confirming that *Ts-hox-3* and the other *hox-3* genes are *Hox3* orthologs.

Fig 4.2. Phylogenetic analysis of *ceh-13*, *hox-3* and *lin-39* genes



B Parsimony

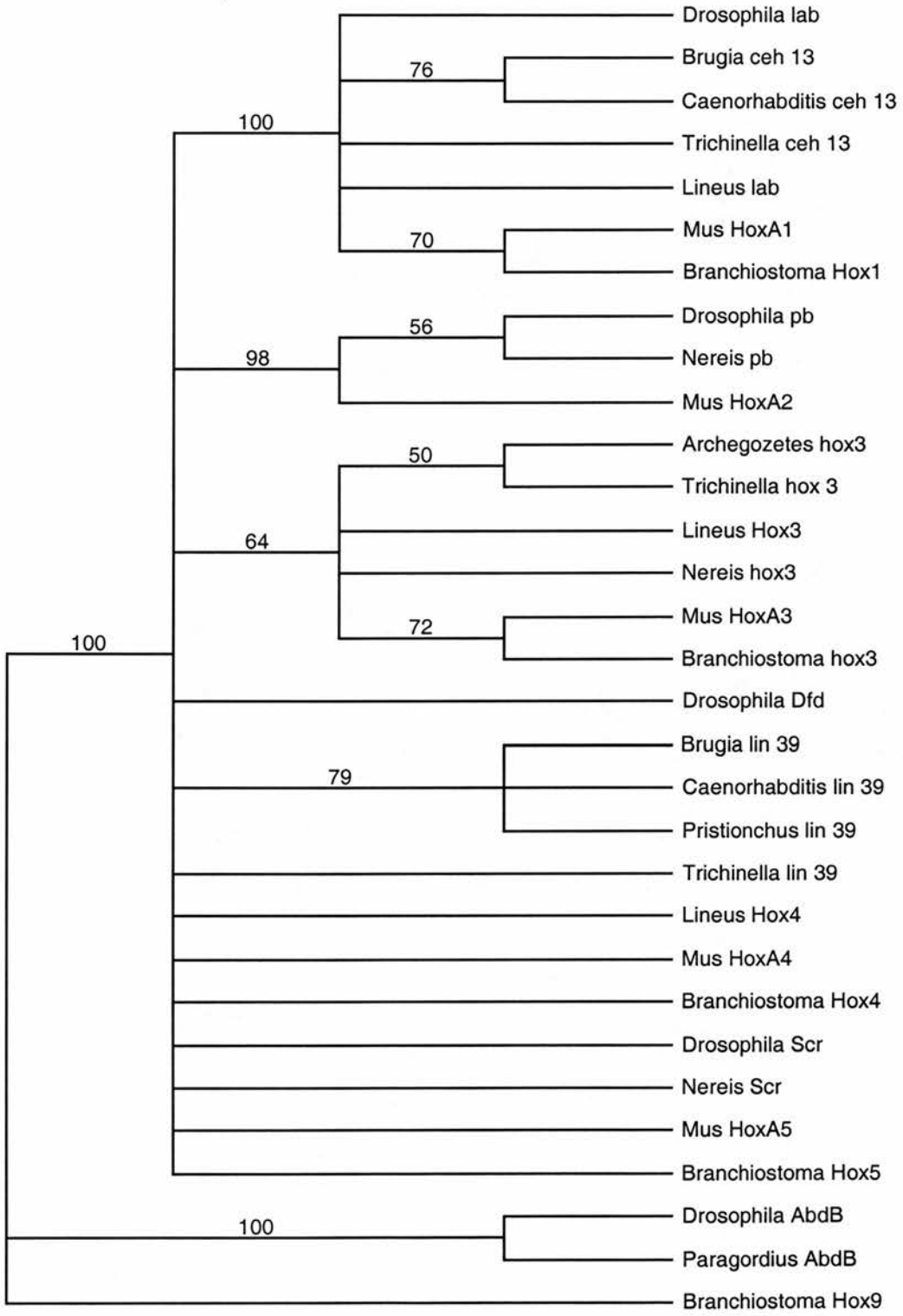

Drosophila lab
Brugia ceh 13
Caenorhabditis ceh 13
Trichinella ceh 13
Lineus lab
Mus HoxA1
Branchiostoma Hox1
Drosophila pb
Nereis pb
Mus HoxA2
Archegozetes hox3
Trichinella hox 3
Lineus Hox3
Nereis hox3
Mus HoxA3
Branchiostoma hox3
Drosophila Dfd
Brugia lin 39
Caenorhabditis lin 39
Pristionchus lin 39
Trichinella lin 39
Lineus Hox4
Mus HoxA4
Branchiostoma Hox4
Drosophila Scr
Nereis Scr
Mus HoxA5
Branchiostoma Hox5
Drosophila AbdB
Paragordius AbdB
Branchiostoma Hox9
76
100
70
56
98
50
64
72
100
79
100

Fig 4.2. Phylogenetic analysis of *ceh-13*, *hox-3* and *lin-39* genes

A Neighbor joining analysis of the relationship between anterior Nematode paralogous groups and those from other phyla. Numbers above nodes are percentage bootstrap support values over 1000 replicates. In this analysis the posterior AbdB genes of of *Drosophila melanogaster* and *Paragordius robusta* (Nematomorph) and the *Hox9* gene of *Branchiostoma Floridae* (Amphioxus) were specified as outgroups. Tree topology supports the inclusion of the *ceh-13*, *hox-3* and *lin-39* nematode gene groups in the Bilaterian paralogous groups of *Hox1*, *Hox3* and *Hox4* respectively. Branch lengths are to scale and represent the mean number of changes per site.

B Maximum Parsimony analysis of the same groups described above also supports the inclusion of the *ceh-13*, *hox-3* and *lin-39* nematode gene groups in the Bilaterian paralogous groups of *Hox1*, *Hox3* and *Hox4* respectively. Numbers above nodes are percentage bootstrap support over 100 replicates.

For the alignment for this analysis and full species names see Appendix 3 (Fig 1 and Table 1)

Fig 4.3 Maximum Parsimony analysis of *Ts-hox-3* with other *Hox3* group genes

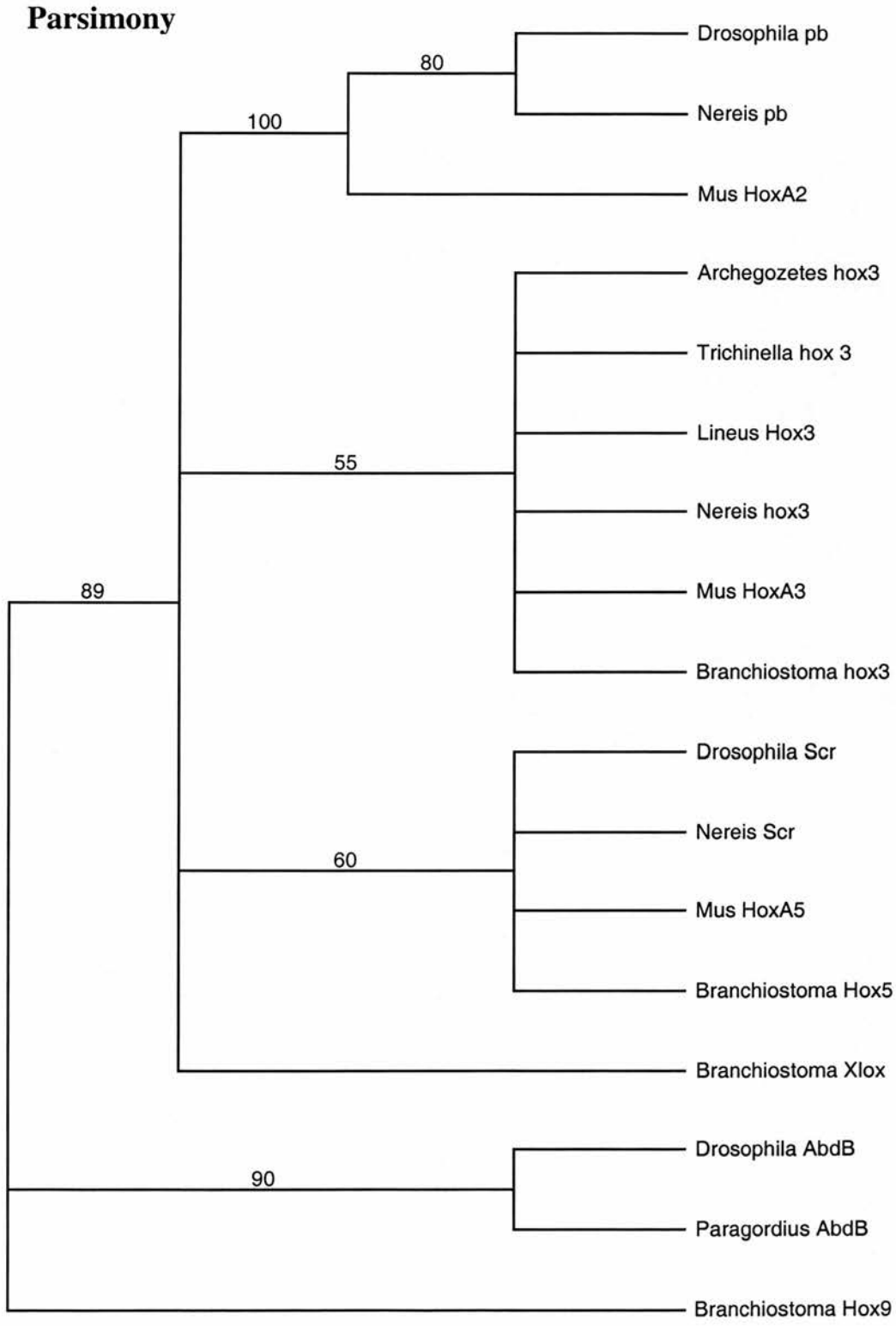


Fig 4.3 Maximum Parsimony analysis of *Ts-hox-3* with other *Hox3* group genes

A phylogenetic tree of *Ts-hox-3* and *Hox3* genes of other phyla showing that *Ts-hox-3* is likely to be a *Hox3* ortholog. The numbers above nodes are percentage bootstrap support values over 100 replicates. The Ts-hox-3 gene groups with the Hox3 genes of other phyla with bootstrap support (>50%). Inclusion of the *Branchiostoma floridae Xlox* gene does not affect the inclusion of *Ts-hox-3* in this clade. Analysis including other ParaHox genes of the *Xlox* group also shows that *Ts-hox-3* is a *Hox3* paralog. *AbdB* genes of of *Drosophila melanogaster* and *Paragordius robusta* (Nematomorph) and the *Hox9* gene of *Branchiostoma floridae* (Amphioxus) were specified as ougroups.

For the alignment for this analysis and full species names see Appendix 3 (Fig 1 and Table 1).

4.4. Phylogenetic analysis of nematode central Hox genes with those of other phyla.

The evolution of central Hox genes across the Bilateria remains uncertain, as sequence evolution has obscured orthology relationships. Within the Ecdysozoa four central groups *ftz*, *Antp*, *Ubx* and *abdA* are recognised (de Rosa et al., 1999). Four central genes are also present in the Lophotrochozoa and are most parsimoniously described as being direct orthologs of those in the Ecdysozoa although the *Lox2* and *Lox4* groups may have arisen through an independent duplication (de Rosa et al., 1999). The Deuterostomata have three central Hox genes corresponding to the *Hox6-8* groups that have not previously been described as being orthologs of any of the protostome central genes. The exact pattern of gene duplication and gene loss that has given rise to the current picture of extant Hox gene distribution is not clear. Often the most parsimonious explanation of the data has been proposed as the most likely. However, the instability observed for the evolution of the nematode Hox genes and Hox cluster (a story of gene loss and rapid sequence evolution) must encourage a closer scrutiny of some of these assumptions. Here the current picture of central Hox gene evolution is reconsidered with the addition of the new central Hox genes from the Nematoda. In this analysis examples of 60 amino acid homeodomains of each of the central genes characteristic of the two major protostome clades and of deuterostomes have been included and analysed with the nematode central genes.

A combined methods approach has been used to assess the relationship of the central nematode Hox genes with those of other phyla. Predictably the small amount of phylogenetic information present in the sequences and their overall high level of similarity results in minimal resolution when bootstrap support (>50%) is set as a requirement for tree topology for NJ and UPGMA analyses (see Fig 4.4 A for NJ and Fig 4.4 B for UPGMA). Overall tree topology when all nematode central Hox genes are included tends to group all the nematode genes together at the base of the tree. This appears to be due to a higher rate of amino acid change that has led to divergent nematode sequences. However, there is bootstrap support for the *mab-5* group and the *As/Bm-ant-1* group by both distance methods. The *Sr-ant-1* gene groups with the *Lox2* genes of the Lophotrochozoa protostome clade by UPGMA analysis. This

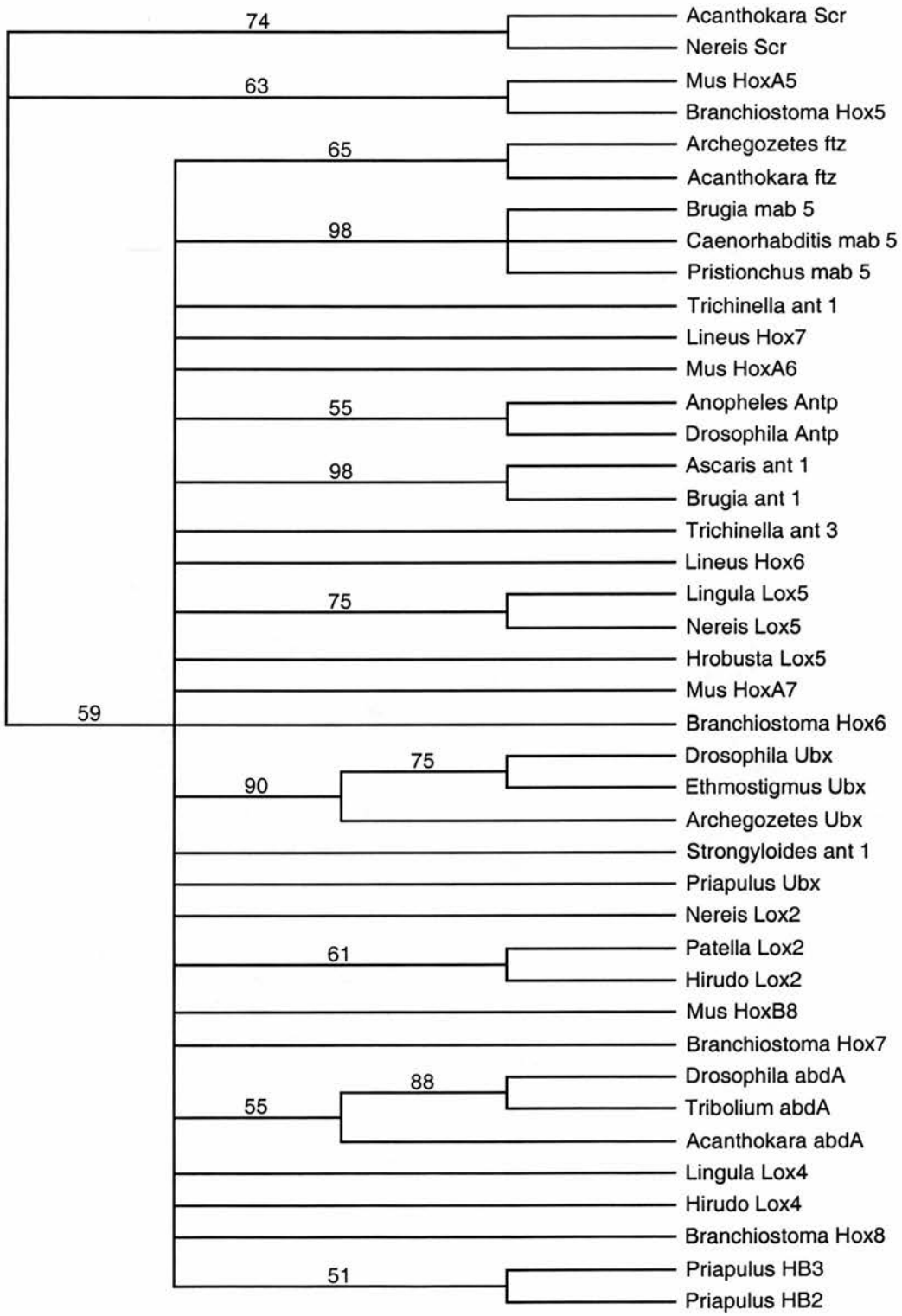
supports the hypothesis that it is actually a contaminant as the association with the *Lox-2* group is so strong.

In parsimony analysis only groups present in the majority concensus of all equally most parsimonious trees were maintained in an overall tree topology (Fig 4.4 C). This analysis places the *Ts-ant-3* gene with the *mab-5* group and the *Ts-ant-1* gene with the *As/Bm-ant-1* group, although the second group was not recovered (<50% support) over a 100 bootstrap replicates (Fig 4.4 C). The *Sr-ant-1* gene groups with the *Lox2* genes and is supported by bootstrap analysis. The *mab-5* group appears to group with the *ftz/Lox5* genes outside the major cluster of central Hox genes containing the *Antp/Ubx/abdA/Lox2/Lox-4/Hox6-8* genes. The *As/Bm-ant-1* group is included within the main cluster of central genes. This suggests that these two NPGs are probably orthologous to two different paralogous groups when compared to other phyla.

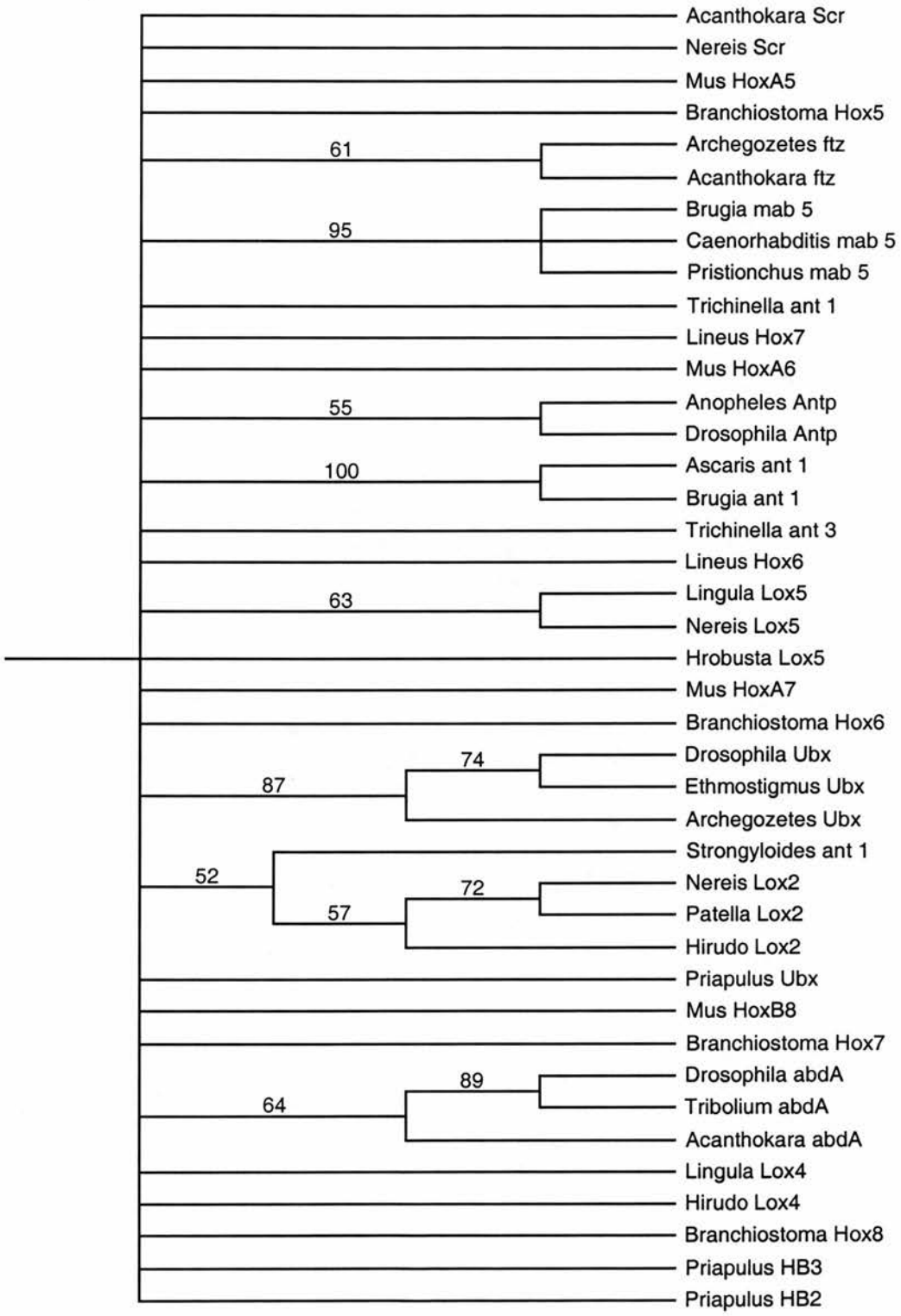
The combined analysis suggests that the nematode central group genes represent at least three different NPGs, the *mab-5* group corresponding to the *Ftz/Lox5* group of protostomes (Telford, 2000a), another central gene group *As/Bm-ant-1* possibly orthologous to the *Antp* gene of protostomes and a third central gene represented by *Sr-ant-1* corresponding to the *Lox2* group previously though to be specific to Lophotrochozoa. The *Ts-ant-3* gene may be a divergent ortholog of the *mab-5* genes of other nematodes and the *Ts-ant-1* gene may be an ortholog of the *As/Bm-ant-1* group. Cloning of central group genes from other clade I nematodes or from the unsampled clade II may clarify this (see Fig 2.1). This analysis agrees with the proposals made in chapter 2 about the orthology arrangements of the central genes (see Fig 2.9).

Figure 4.4. Phylogenetic analysis of central nematode Hox genes

A Neighbour joining



B UPGMA


Acanthokara Scr
Nereis Scr
Mus HoxA5
Branchiostoma Hox5
61
Archegozetes ftz
Acanthokara ftz
Brugia mab 5
95
Caenorhabditis mab 5
Pristionchus mab 5
Trichinella ant 1
Lineus Hox7
Mus HoxA6
55
Anopheles Antp
Drosophila Antp
100
Ascaris ant 1
Brugia ant 1
Trichinella ant 3
Lineus Hox6
63
Lingula Lox5
Nereis Lox5
Hrobusta Lox5
Mus HoxA7
Branchiostoma Hox6
74
Drosophila Ubx
87
Ethmostigmus Ubx
Archegozetes Ubx
Strongyloides ant 1
52
72
Nereis Lox2
57
Patella Lox2
Hirudo Lox2
Priapulus Ubx
Mus HoxB8
Branchiostoma Hox7
89
Drosophila abdA
64
Tribolium abdA
Acanthokara abdA
Lingula Lox4
Hirudo Lox4
Branchiostoma Hox8
Priapulus HB3
Priapulus HB2

C Parsimony

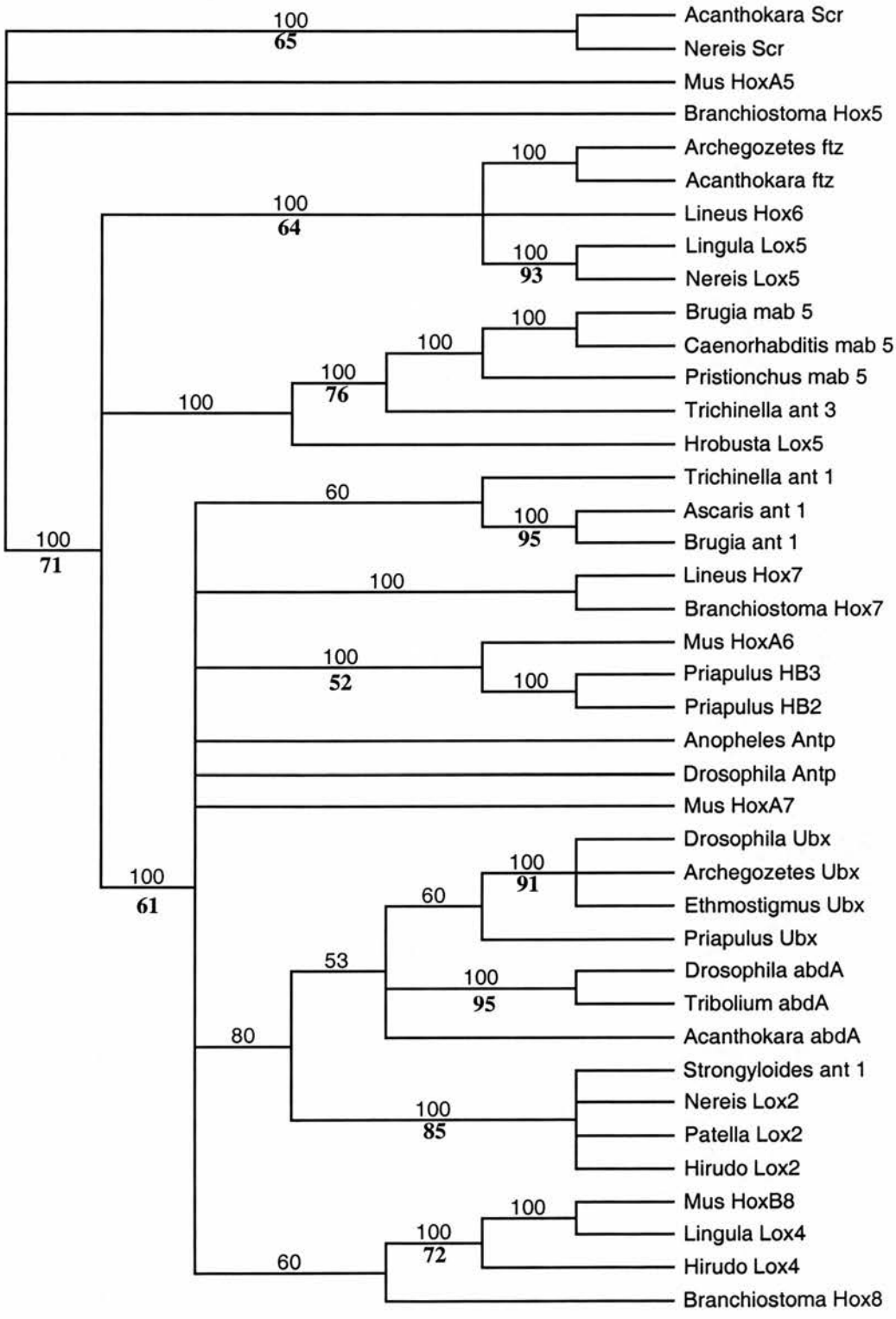

Acanthokara Scr
Nereis Scr
Mus HoxA5
Branchiostoma Hox5
Archegozetes ftz
Acanthokara ftz
Lineus Hox6
Lingula Lox5
Nereis Lox5
Brugia mab 5
Caenorhabditis mab 5
Pristionchus mab 5
Trichinella ant 3
Hrobusta Lox5
Trichinella ant 1
Ascaris ant 1
Brugia ant 1
Lineus Hox7
Branchiostoma Hox7
Mus HoxA6
Priapulus HB3
Priapulus HB2
Anopheles Antp
Drosophila Antp
Mus HoxA7
Drosophila Ubx
Archegozetes Ubx
Ethmostigmus Ubx
Priapulus Ubx
Drosophila abdA
Tribolium abdA
Acanthokara abdA
Strongyloides ant 1
Nereis Lox2
Patella Lox2
Hirudo Lox2
Mus HoxB8
Lingula Lox4
Hirudo Lox4
Branchiostoma Hox8

Figure 4.4. Phylogenetic analysis of central nematode Hox genes

A Neighbour joining analysis of selected central Bilaterian Hox genes. Numbers above nodes are percentage bootstrap support over 1000 replicates. None of the nematode central genes are grouped with those of other phyla with bootstrap support (>50%). In addition the *Ts-ant-1*, *Ts-ant-3* and *Sr-ant-1* genes do not group with either the *As/Bm-ant-1* or *mab-5* NPGs **B** UPGMA analysis of selected central Bilaterian Hox genes. Numbers above nodes are percentage bootstrap support over 1,000 replicates. The *Sr-ant-1* gene groups with the *Lox2* genes of the Lophotrochozoa. **C** Parsimony analysis of selected central Bilaterian Hox genes. Numbers above nodes represent the percentage of 10,000 most parsimonious trees in which that node was represented. Numbers below nodes are percentage bootstrap replicates over 100 bootstrap replicates with MAXTREES set to 10,000. The *Sr-ant-1* gene again groups with the *Lox2* genes of the Lophotrochozoa. The *Ts-ant-3* gene groups with the *mab-5* group with bootstrap support that in turn groups with the *ftz/Lox5* group genes from other Bilaterians. The *Bm/As-ant-1* genes are grouped with the rest of the central genes, suggesting that they represent a separate paralogous group from the *mab-5* genes. The central genes form a separate group from the *Ftx/Lox5* genes with bootstrap support (>50%).In all analyses the *Hox5* group genes are specified as an outgroup. The alignment for this analysis is presented in Appendix 3 (Fig 1) along with full species names (Appendix 3, Table 1).

4.5. Phylogenetic analysis of the nematode *egl-5* and *php-3* posterior group genes

There are two posterior group genes present throughout the Nematoda. The *php-3* group is clearly an ortholog of the *AbdB* orthology group so far found only in members of the Ecdysozoa (de Rosa et al., 1999). Characteristic residue analysis of the *egl-5* group shows that it has diverged considerably from other Hox genes but it retains some of the residues characteristic of the *AbdB* orthology group (see chapter 2).

Phylogenetic analysis of posterior group genes shows that the *php-3* group is related to the posterior genes of other Ecdysozoa (Fig 4.5) forming a distinct clade from the *Post1* and *Post2* groups of the Lophotrochozoa and the Deuterostome posterior genes. Analysis of the *egl-5* group confirms that they are not related to any other group of posterior genes. The cloning of more posterior Hox genes from other Ecdysozoan phyla may shed light on whether this gene is a divergent posterior gene that has arisen only in the Nematoda or whether it arose earlier in Bilaterian evolution. The observation that *Ts-egl-5* is just as divergent as all the other nematode *egl-5* genes and has all the characteristic residues implies that the *egl-5* gene had already diverged at the root of the Nematoda. Perhaps the best clue to the origin of the *egl-5* group is the maintenance of some residues characteristic of the *AbdB* group (see chapter 2).

Fig 4.5 Phylogenetic analysis of the nematode *php-3* group

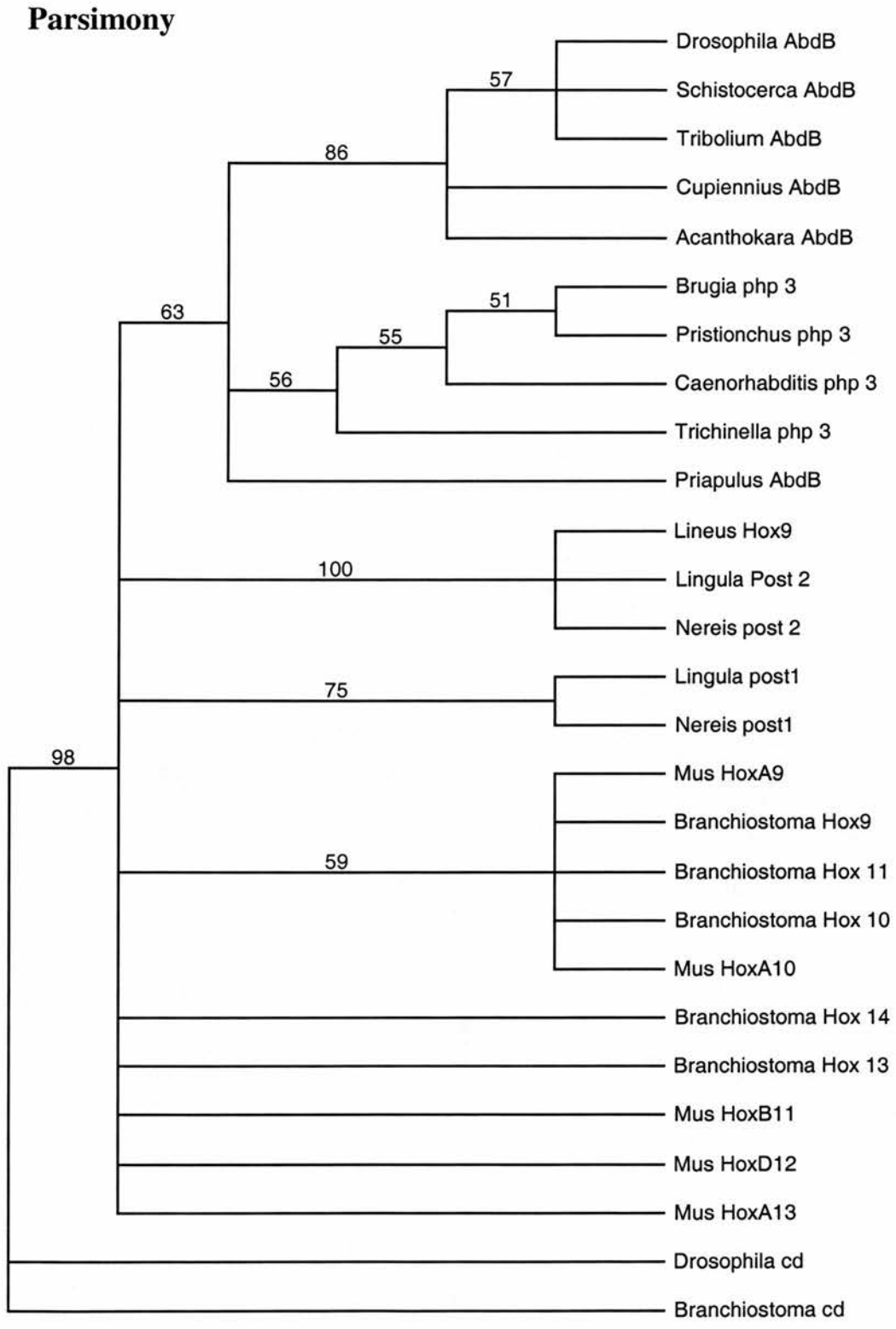


Fig 4.5 Phylogenetic analysis of the nematode *php-3* group

A phylogenetic tree of posterior group genes through the Bilateria with caudal group genes as outgroups. Numbers above nodes represent percentage bootstrap support over 100 replicates. The *php-3* group genes of the Nematoda group with the *AbdB* group genes of the other Ecdysozoan phyla (Arthropoda Priapulida Onycohphora) with bootstrap support (>50%). The *Post1* and *Post2* genes of the Lophotrochozoa are also recovered with bootstrap support.

The alignment for this analysis is presented in Appendix 3 (Fig 1) along with full species names (Appendix 3, Table 1).

4.6. Assessing the evolutionary rate of Nematode Hox genes

The robust molecular phylogeny available for the Nematoda means that we can be sure of the relationships of the taxa from which Hox genes are available. Superimposing the Hox gene sequences over this phylogeny allows the rate of amino acid substitutions in the Nematoda to be accurately assessed. One observation that can be made by simple alignment of some NPGs is that despite the presence of some ancestral residues when compared to other phyla and characteristic nematode residues there is a lot of variation in homeodomain sequences. This and the fact that nematode genes are divergent from their orthologs in other phyla suggests that Nematode Hox genes may be evolving more rapidly then those of other phyla. By specifying the relationships between the orthologous genes across the Bilateria we can assess the number of unambiguous amino acid changes within their homeodomains. This analysis has been conducted on the *ceh-13* group, the *hox-3* group, the *lin-39* group and *php-3* group.

Specifying the relationship of the nematode genes in each NPG and comparing them by parsimony analysis with genes in the same paralogous groups from other phyla, the rate of evolution in the Nematoda can be measured. One problem with this approach is that many sequences do not span the full 60 amino acid homeodomain. If the total number of amino acid substitutions is estimated by parsimony and taken to represent the rate of evolution of taxa, then the rate wherever full-length homeodomains are not known will be underestimated. This will be accentuated where the unknown regions include the most variable residues within a paralogous group. However, in the case of some of the NPGs membership is across the whole phyla and can be compared with the rest of the Bilateria. This analysis will also identify taxa that are evolving more rapidly than other members of their phyla. Such cases may indicate a change in function of the Hox gene in question (as with hindsight in the case of the *Drosophila zen* and *ftz* genes for example (Falciani et al., 1996; Lohr et al., 2001; Telford and Thomas, 1998)) or possibly that the gene is not the true ortholog from the species in question but perhaps a divergent recent paralog of it. For this analysis a phylogenetic tree for the Bilateria was assumed, placing the Nematoda with Arthropoda and other Ecdysozoa with members of the

Lophotrochozoa and Deuterostomata as outgroups to this clade (see chapter 1). Within the Arthropoda the Hexapoda were allied with the Crustacea as confirmed by a number of recent molecular phylogenetic studies (Blaxter, 2001; Giribet et al., 2001; Hwang et al., 2001). No assumptions were made about the relationship of different Lophotrochozoan taxa that are not as well sampled as the phylum Arthropoda with respect to Hox gene content.

By this analysis it is clear that the *ceh-13* group of the Nematoda has evolved rapidly in comparison to the *Hox1* group genes of other phyla (Fig 4.5 A) with the genes of clade IV and V evolving the fastest. For example by this simplistic analysis a total of at least 13 amino acid substitution events in the known region of the homeodomain *Mj-ceh-13* genes may have occurred since the last common ancestor of the Bilateria. Of these 12 have occurred since the divergence of the Nematode lineage from the Arthropods and Priapulids. This immediately suggests a loss of constraint on sequence evolution in the homeodomain of this gene in the Nematode lineage. It will be fascinating to find out what evolutionary changes in developmental mechanisms at the molecular level have facilitated this. For example it may be possible to test by using *C. elegans* as a heterologous transgenic system how much sequence evolution will be tolerated by *ceh-13* homeodomain before normal function of the gene is perturbed. Is the sequence evolution that is observed in the *ceh-13* homeodomain occurring to maintain specificity for downstream targets in the different species or are the changes tolerated as long as they occur within certain limits that do not perturb function because the amount of specificity required is minimal?

The *Hox3* genes of insects are assumed have undergone rapid evolution as they have lost their embryonic function (Akam, 1998b; Falciani et al., 1996; Telford and Thomas, 1998). It has been suggested that the divergent sequence evolution has occurred as a loss of selective constraint on the function of the homeodomain (Akam, 1998b; Falciani et al., 1996; Telford and Thomas, 1998). Rapid sequence evolution is also observed for the nematode *hox-3* group across the whole phylum, again highest in clades IV and III after the divergence form the more basal nematode *T. spiralis* (Fig 4.5 B). The *hox-3* gene has definitely been lost from the *C. elegans* lineage and whatever function it serves in other nematodes it is not required for development in

this species. Thus coupled with loss of a *Hox2* gene the rapid evolution of the *ceh-13* and *hox-3* groups suggests that importance of Hox genes in the specification of the anterior of the nematode body plan is fundamentally different from other phyla.

The divergent *lin-39* group has also evolved at a higher rate than other Ecdysozoa across the phylum and again many of the unambiguous amino acid substitutions have occurred since the divergence of clades III, IV and V from the more basal *T. spiralis*. The amount of evolution of the two Lophotrochozoa *Hox4* genes is also high but cannot be interpreted as being across the whole group as to just two examples from separate phyla are included. The *Hox4* genes across the Arthropoda are relatively highly conserved. In clade V nematodes *C. elegans* has a central role in patterning the vulval precursor cells, but despite its apparently conserved function in this clade it is obvious that significant changes have occurred in how the gene controls and is itself regulated in this process (Grandien and Sommer, 2001; Sommer et al., 1998). It appears that these changes are not due to changes in the coding sequence of this gene but are more likely to be in the upstream regulatory regions. In this single known example at least, it is clear that sequence evolution is not linked to an observed functional change, although the divergence between *P. pacificus* and *C. elegans* is relatively not that great (Fig 4.5 C).

The posterior genes of Deuterostomata have been shown to be evolving more rapidly than other paralogous group (*laxitas terminalis*) and to an extent that they cannot be resolved between different taxonomic groups (posterior flexibility) (Ferrier et al., 2000; van der Hoeven et al., 1996). The multiple posterior genes of the Deuterstoma have not been included here; only the *Hox9* genes of Amphioxus and mouse have been included as likely orthologs of each other (although they still do not resolve phylogenetically). Instead the *php-3* group is compared to the other available Ecdysozoa *AbdB* group genes and the posterior genes of the Lophotrochozoa. The pattern of evolution is again similar to that of the other genes tested so far if less pronounced than for the anterior genes. A higher rate of amino acid substitution is observed across nematode clades V, IV and III. The posterior gene of the Arthropods is highly conserved and little change is scene. In contrast the posterior group genes of the Lophotrochozoa, *Post1* and *Post2* appear to evolving

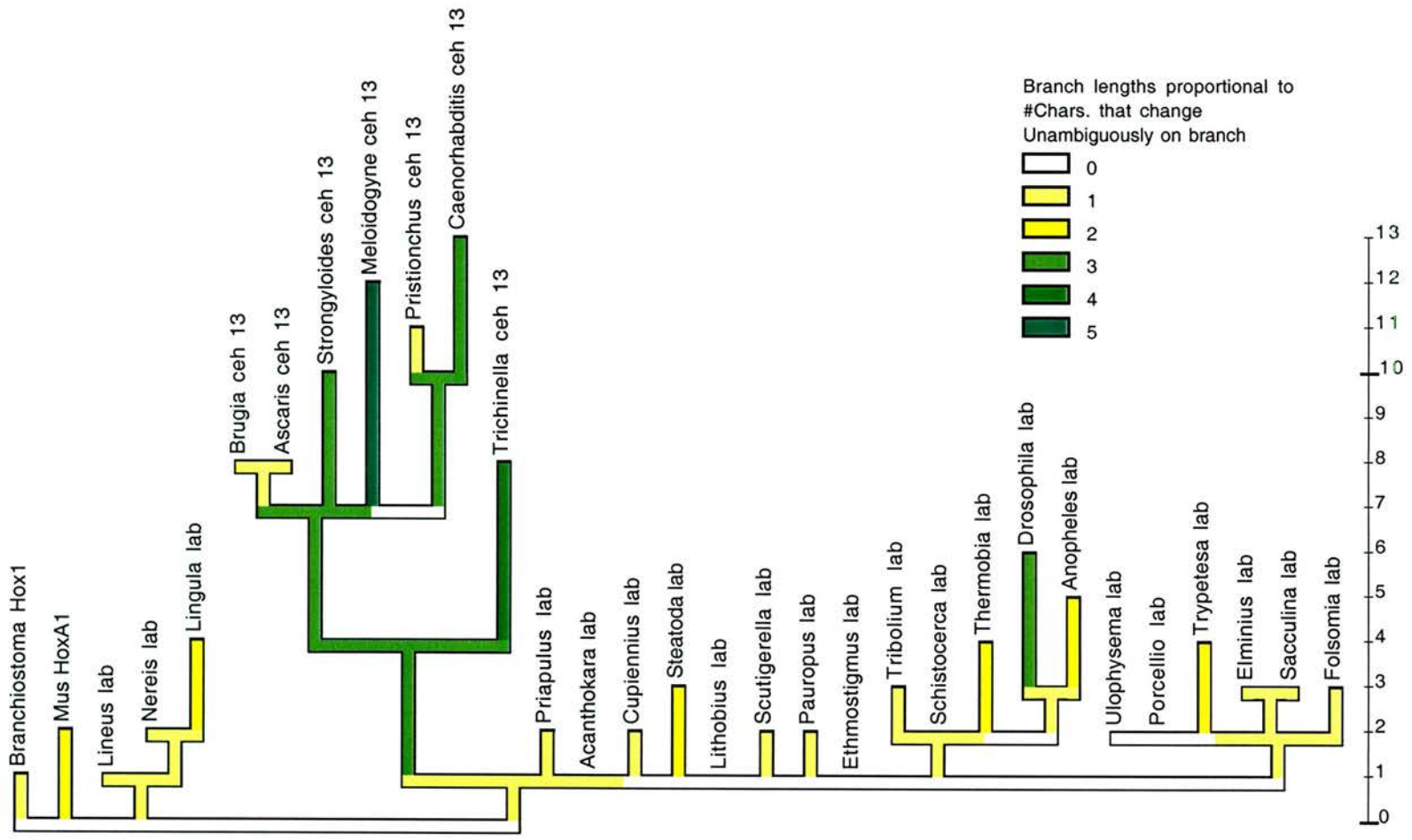
more rapidly. This is particularly true if we assume that the gene duplication giving rise to two posterior group genes occurred within the Lophotrochozoan lineage.

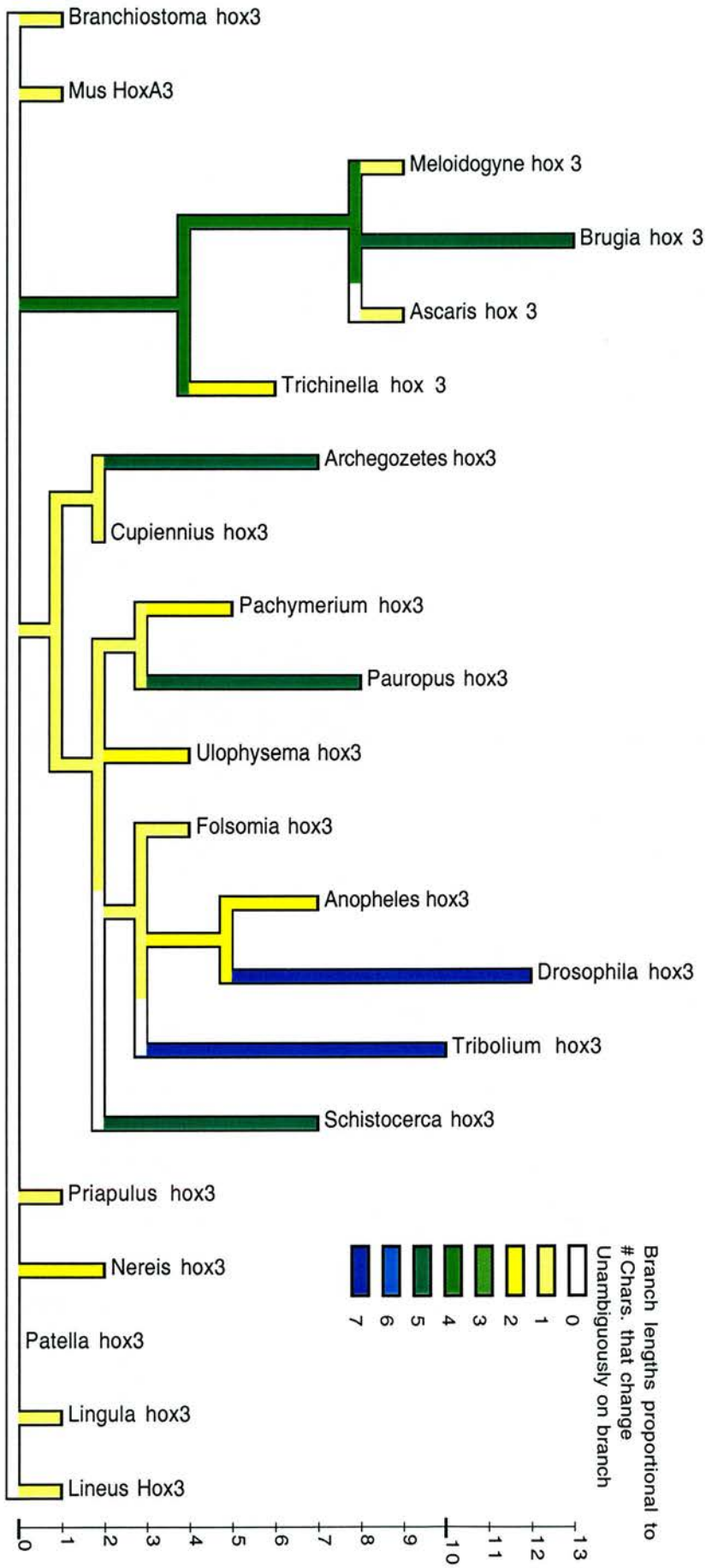
Within the Nematoda the *S. ratti php-3* gene is evolving very rapidly. As mentioned earlier this may be an indication that it is not the true ortholog of the other *php-3* genes. Reanalysis of its alignment with the other nematode genes shows that it is clearly divergent from the *php-3* group, it contains only one characteristic nematode residue and one residue characteristic of the Ecdysozoan *AbdB* group. However, it is not closer in sequence to the *egl-5* group by phylogenetic analysis and does not bear similarity to the third posterior gene in *C. elegans nob-1*. Possibly it is the product of a recent duplication event within the *S. ratti* lineage. The extra sequence cloned by sequencing of a truncated 5' RACE product confirmed that this gene was a posterior class gene as it retains characteristic residues of this group (see chapter 2).

Overall nematode Hox genes are evolving more rapidly across the whole of the phylum than those of the Arthropoda and probably then the rest of Bilateria. The central genes are certainly evolving rapidly making it difficult to assign them to central gene paralogy groups in other phyla. The analysis here shows that the other NPGs are also evolving rapidly. Some of this evolution appears to have occurred before the divergence of the different nematode lineages but much of it is confined to clades III, IV and V. It will be interesting to see whether the apparent increase in amino-acid substitutions is a sign of functional changes on a gene by basis or reflects a general loss of evolutionary constraint on Hox genes due to the deterministic nature of nematode development.

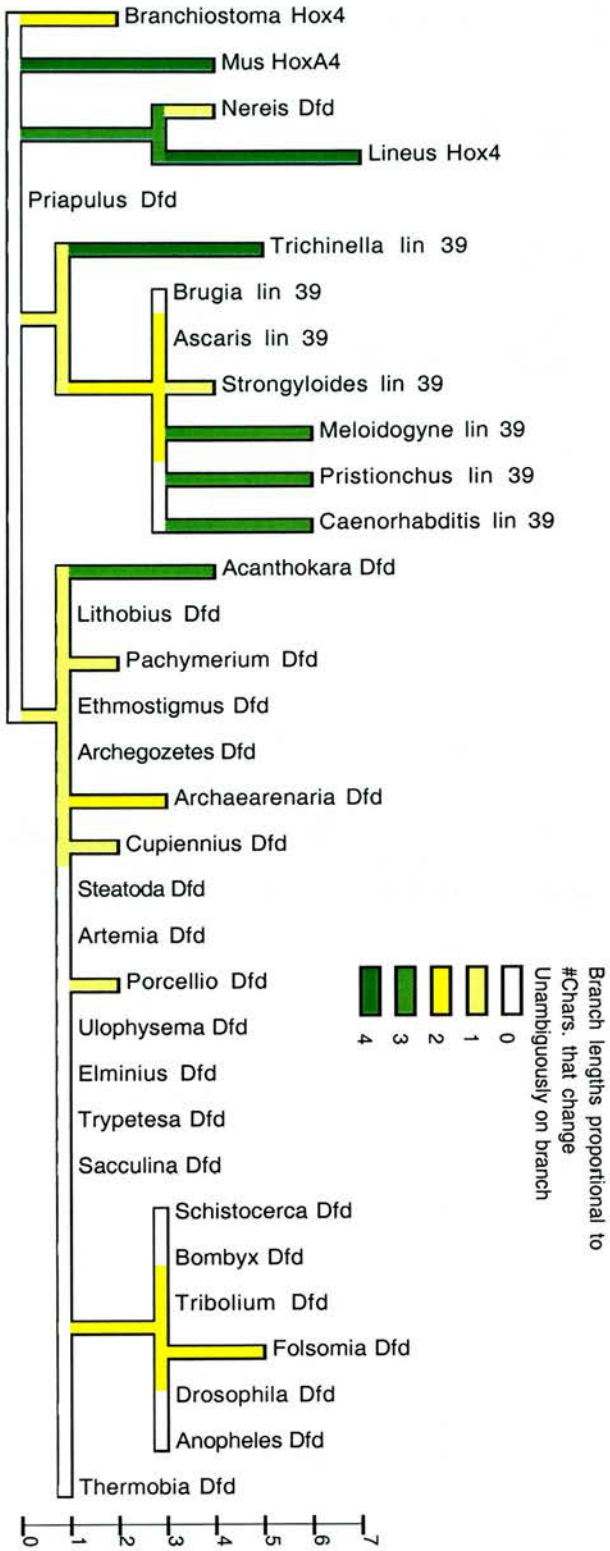
Fig 4.5. Comparison of the evolutionary rate of Hox genes from nematodes with those other phyla

A




Branchiostoma hox3
Mus HoxA3
Meloidogyne hox 3
Brugia hox 3
Ascaris hox 3
Trichinella hox 3
Archegozetes hox3
Cupiennius hox3
Pachymerium hox3
Pauropus hox3
Ulophysema hox3
Folsomia hox3
Anopheles hox3
Drosophila hox3
Tribolium hox3
Schistocerca hox3
Priapulus hox3
Nereis hox3
Patella hox3
Lingula hox3
Lineus Hox3
0 1 2 3 4 5 6 7 8 9 10 11 12 13
7 6 5 4 3 2 1 0
Branch lengths proportional to
Chars. that change
Unambiguously on branch

B


Branchiostoma Hox4
Mus HoxA4
Nereis Dfd
Lineus Hox4
Priapulus Dfd
Trichinella lin 39
Brugia lin 39
Ascaris lin 39
Strongyloides lin 39
Meloidogyne lin 39
Pristionchus lin 39
Caenorhabditis lin 39
Acanthokara Dfd
Lithobius Dfd
Pachymerium Dfd
Ethmostigmus Dfd
Archegozetes Dfd
Archaearenaria Dfd
Cupiennius Dfd
Steatoda Dfd
Artemia Dfd
Porcellio Dfd
Ulophysema Dfd
Elminius Dfd
Trypetesa Dfd
Sacculina Dfd
Schistocerca Dfd
Bombyx Dfd
Tribolium Dfd
Folsomia Dfd
Drosophila Dfd
Anopheles Dfd
Thermobia Dfd
0 1 2 3 4 5 6 7
4 3 2 1 0
Branch lengths proportional to
#Chars. that change
Unambiguously on branch

C

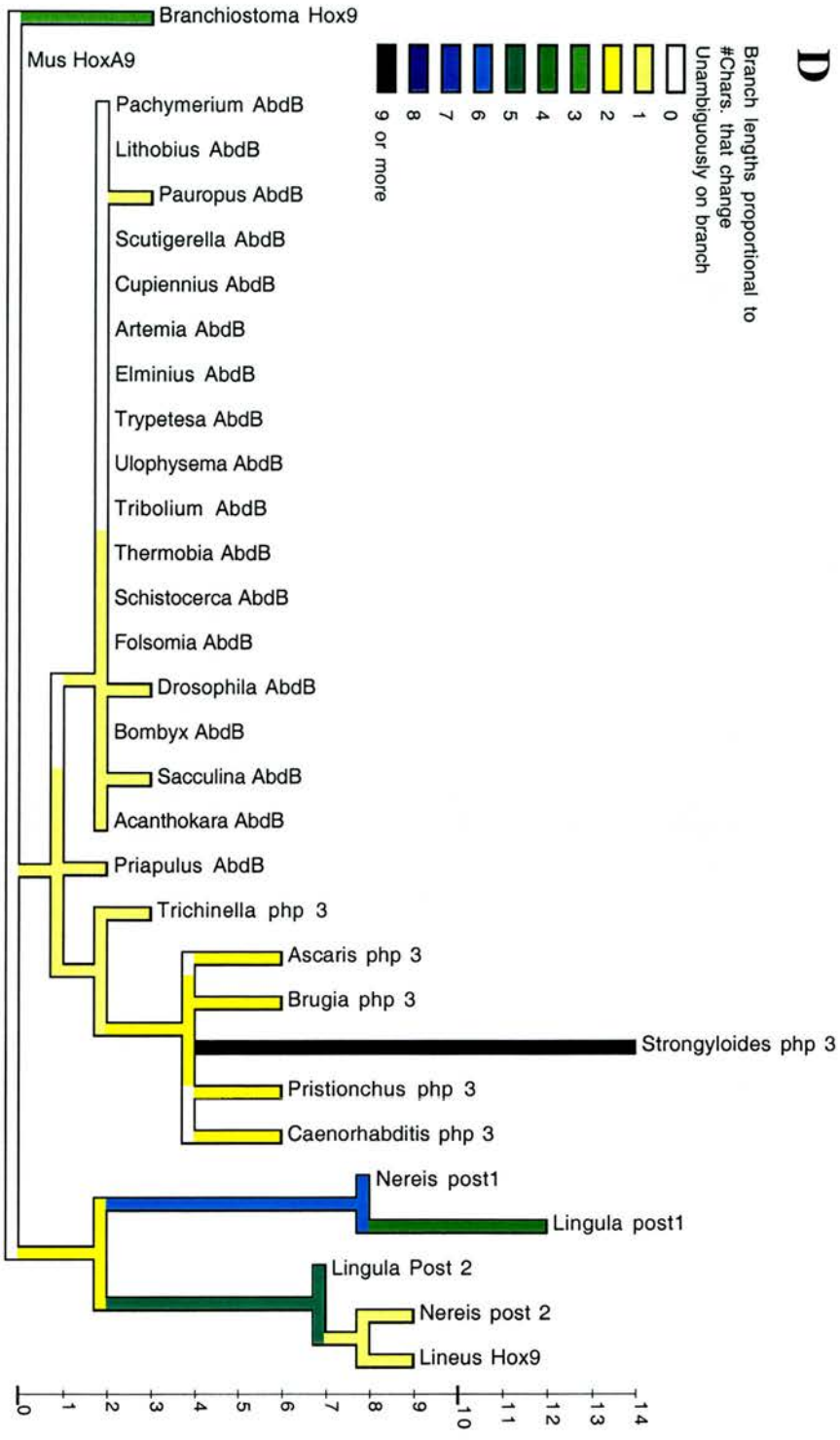

D
Branch lengths proportional to
#Chars. that change
Unambiguously on branch
0
1
2
3
4
5
6
7
8
9 or more
Branchiostoma Hox9
Mus HoxA9
Pachymerium AbdB
Lithobius AbdB
Pauropus AbdB
Scutigerella AbdB
Cupiennius AbdB
Artemia AbdB
Elminius AbdB
Trypetesa AbdB
Ulophysema AbdB
Tribolium AbdB
Thermobia AbdB
Schistocerca AbdB
Folsomia AbdB
Drosophila AbdB
Bombyx AbdB
Sacculina AbdB
Acanthokara AbdB
Priapulus AbdB
Trichinella php 3
Ascaris php 3
Brugia php 3
Strongyloides php 3
Pristionchus php 3
Caenorhabditis php 3
Nereis post1
Lingula post1
Lingula Post 2
Nereis post 2
Lineus Hox9
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14

Fig 4.5. Comparison of the evolutionary rate of Hox genes from nematodes with those from other phyla

An assumed phylogeny based on the interrelationship of the nematode species (see chapter 2), the relationship of the Nematoda to the other phyla as described in chapter 1 was constructed MacClade 4.02. The relationship of taxa in other parts of the tree were specified according to the separation of protostomes into two major clades the Lophotrocozoa and the Ecdysozoa. Within the Arthropoda the Crustacea were grouped with the Hexapoda as suggested by recent molecular phylogenetic analysis. MacClade 4.02 was used to trace all unambiguous amino acid changes in sets of orthologous genes across the Bilateria within the pre-specified phylogeny. Changes were measured based on a parsimony analysis that assumed the most parsimonious ancestral sequence for the Bilateria based on the included sequences. **A** A comparison of unambiguous amino acid changes in *Hox*1 group genes across the Bilateria. **B** A comparison of unambiguous amino acid changes in *Hox3* group genes across the Bilateria. **C** A comparison of unambiguous amino acid changes in *Hox4* group genes across the Bilateria. **D** A comparison of unambiguous amino acid changes in posterior group genes across the protostome phyla.

The alignment on which this analysis was based are taken form the major alignment presented in Appendix 1 along with full species names (Appendix 1, Table 1) of the taxa included.

4.7. Overview of Hox gene evolution in the Nematoda

The current picture of Hox gene evolution in the Nematoda is an intriguing one and offers the possibility of investigating some processes in the evolution of developmental mechanisms that are not well understood. Firstly we can unequivocally say that at of the six Hox genes that have been lost in the *C. elegans* lineage by comparison with other protostome phyla at least three have been lost within the nematode lineage. While all other protostome taxa studied to date (with the exception of Cirripedia (Mouchel-Vielh et al., 1998)) have maintained a full complement of Hox genes. In some taxa some genes may have lost their homeotic function and assumed an alternate function. Such changes have been characterised in insects where both the *Hox3* ortholog zen genes and the central *ftz* class genes have assumed new non-homeotic functions (Falciani et al., 1996; Lohr et al., 2001; Telford, 2000b; Telford and Thomas, 1998). These functional changes have been accompanied by rapid sequence evolution away from their orthologs in other phyla. This rapid sequence evolution is also a characteristic of the nematode Hox genes. The anterior class Hox gene *ceh-13* group and the *hox-3* group are evolving more rapidly then across the whole phylum then their orthologs in the rest of the Bilateria. Two of the three central Hox gene groups represented in the Nematoda also appear to be evolving rapidly as they are distant from their orthologs in other phyla.

4.7.1. The tempo of Hox gene evolution in the Nematoda

The phylogenetic analysis in this chapter confirms the presence of 7 Bilaterian paralogy groups in the Nematoda represented by eight different gene groups. Of these eight groups 7 have multiple representatives. Attempts to clone the full homeodomain and flanking regions of this gene will help to confirm whether this gene is a contaminating *Lox2* gene from a lophotrochozoan or possibly a real nematode gene. The central protostome genes in the *Ubx*, *Lox2*, *abdA*, *Lox4* groups have some conserved characteristic residues at both ends of the homeodomain (de Rosa et al., 1999). Nematode genes where flanking residues are known do not have

the any of the characteristic flanking residues seen in the other protostome groups. For example the *lin-39* group has not maintained the characteristic LPTNKXNXR motif conserved across *Dfd/Hox4* orthologs of the Bilateria (de Rosa et al., 1999). Presumably for this peptide motif to be maintained it must have some conserved ancestral functional significance for molecular interactions of the *Dfd/Hox4* proteins. Its loss from the Nematoda immediately suggests that this interaction is not conserved in this phylum.

The loss of characteristic flanking residues is one measure of the relatively high rate of Nematode Hox gene evolution. Another is the rate of amino acid substitution observed within the Nematoda that is very high compared to other Bilaterian clades. The Arthopoda represent the richest source of data for comparison at the phylum level and the evolution of Nematode Hox genes within the phylum, as estimated by parsimony on an assumed phylogeny, are much faster than that for the Arthropods for the orthology groups tested. The difficulty of assigning some of the nematode central genes to orthology groups is also indicative of a higher rate of Nematode Hox gene evolution.

The confirmation in the analysis in this chapter that nematode Hox genes are evolving rapidly in comparison to other phyla allows the construction of slightly more refined hypothesise about the nature of the evolutionary changes to gene function involved.

4.7.2. Explaining Hox gene loss in the Nematoda, an opportunity to study a special case in the evolution of developmental processes.

The evolution of developmental mechanisms in the Nematoda has led to the loss of six Hox orthology groups from the *C. elegans* lineage, with three of the losses occurring after the radiation of clades III, IV and V of the Phylum (see Fig 2.9, chapter 2). Along with the observed increased rate of sequence evolution this suggests that the role of Hox genes in patterning the anterior posterior axis has changed significantly in nematodes, in ways distinct form those observed from other phyla. Firstly, all the Hox genes appear to have been affected whereas in the other

taxa some Hox genes have an increased rate linked to changes in function (*zen* and *ftz* of *Drosophila*, see above). Secondly Hox genes have been lost implying that they were no longer needed for patterning the nematode body axis, changes in nematode development having obviated a need for them. Ironically the model nematode *C. elegans* may be the most extreme example of this. It is possible that these two observations are linked and various testable hypotheses can be constructed to explain what mechanisms can account for the changes observed.

Considering the sequence conservation seen in the Bilateria for Hox genes in the homeodomain at least, it is possible that conserved ancestral Hox gene interactions (such as that with PBX gene family) have been maintained across the Bilateria. Other interactions may be quite specific to different Bilaterian groups relying on regions without sequence conservation between groups outside of the homeodomain. If the role of nematode Hox genes has been marginalized by the employment of cell lineage mechanisms of development at the level at which Hox genes are ancestrally required for specifiying positional identities (assuming this is the ancestral state), this may introduce some redundancy in their function. We can see three different extant roles for Hox genes in this scenario that are directly observed in *C. elegans*. Firstly, as the nematode development has become more cell lineage dependent some Hox genes are lost altogether from some lineages, this would be the cause of gene loss observed in *C. elegans*. Alternatively some Hox genes may lose their role in embryonic patterning but be co-opted to other roles, maintaining their ancestral boundaries of expression. In this case we might expect significant sequence change, as Hox genes evolved to interact with new partners in different lineages, in both the coding and regulatory regions of genes. The *mab-5*, *lin-39* and *egl-5* genes all appear to be at this point with respect to their roles in *C. elegans*. Thirdly, some Hox genes may be co-opted into a cell lineage mechanism of development with a role in embryogenesis. In this case we would expect the expression of Hox genes to be of cell lineage dependent and independent of position along the body axis. The *ceh-13* and *php-3/nob-1* genes (that appear to function in tandem at the posterior of the embryo) are illustrative of this state, although their expression pattern has not been shown to be cell lineage specific as it has been for *lin-39*. These three possibilities obviously form a framework within which to look at

the function of Hox genes from other nematodes (see chapter 3). By investigating the expression pattern and possibly function of Hox genes in other nematodes it should be possible to see whether the states observed in *C. elegans* are reflected across the Nematoda. These studies would also require some knowledge of the extent to which a cell lineage mode of development is employed in these species (Voronov and Panchin, 1998; Voronov et al., 1998). The processes leading to redundancy and gene loss may include the other two stages observed for *C. elegans* genes, of modified function within a changed developmental mode characterised by integration into the new mode (cell lineage dependent expression not position) followed by the loss of roles in earlier developmental stages as these functions became redundant. Analysis of the Hox genes present in other Nematoda missing from *C. elegans* may provide more detailed insight in to the steps in the evolution of developmental mechanisms that lead to loss of highly conserved genes.

CHAPTER 5: DEVELOPMENT OF RNAi IN THE PARASITIC NEMATODE *BRUGIA MALAYI*; A POTENTIAL METHOD FOR DRUG TARGET SCREENING IN FILARIAL NEMATODES.

5.1. An Overview of dsRNA mediated gene silencing

Perhaps the most fundamental and useful insight into the basic mechanisms of cellular metabolism of the last decade has been the discovery of dsRNA mediated gene silencing in eukaryotes (Fire, 1999; Fire et al., 1998). When double stranded RNA (dsRNA) corresponding to the sense and antisense sequence of an endogenous mRNA is introduced into a cell the cognate mRNA is degraded, resulting in gene silencing (Bass, 2000; Fire et al., 1998). This type of post-transcriptional gene silencing (PTGS), referred to as RNA interference (RNAi), was first observed in *C. elegans* and has now been demonstrated in all five eukaryotic kingdoms (Brown et al., 1999; Catalanotto et al., 2000; Cogoni and Macino, 2000; Elbashir et al., 2001a; Hope, 2001; Li et al., 2000; Maine, 2000; Schoppmeier and Damen, 2001; Shi et al., 2000; Wianny and Zernicka-Goetz, 2000). RNAi in many ways resembles the PTGS that can be observed when a transgene is introduced into a cell (co-suppression) and has overlapping genetic requirements (Ketting and Plasterk, 2000).

Current evidence strongly supports the hypothesis that RNAi is a generic genome surveillance mechanism evolved to combat the deleterious effects of both RNA viruses and transposons (Tabara et al., 1999), both of which may give rise to dsRNA intermediates during their replication. It is likely that the balance between transposon activities and RNAi have had an important role in shaping the structure of the genomes of most organisms (Sharp, 1999). Indeed, at least two plant viruses are known to encode suppressors of dsRNA mediated gene silencing (Anandalakshmi et al., 1998; Voinnet, 2001; Voinnet et al., 2000). It is also apparent that genes related to those responsible for RNAi has been recruited to regulate the timing of some developmental processes by blocking the translation of endogenous mRNA species (Grishok et al., 2001; Grosshans and Slack, 2002; Lau et al., 2001; Lee and Ambros, 2001).

The discovery of RNAi is having two major impacts on molecular biology. Firstly it has opened up a new area of research into a novel, conserved process of fundamental importance to eukaryotic biology (Bass, 2000; Carthew, 2001). Secondly, as a technology, it has offered the potential for performing rapid reverse genetic screens (Bargmann, 2001; Barstead, 2001; Fraser et al., 2000; Gonczy et al., 2000; Maeda et al., 2001) and has possible therapeutic applications (Hope, 2001). This has had and will continue to have a profound effect on research on the major model organisms, but of potentially greater importance is its use on intractable non-model or non-genetic organisms, previously impervious to molecular genetic studies of gene function (Kuwabara and Coulson, 2000a).

Insight into the mechanisms of RNAi has come from detailed genetic and biochemical studies in systems specifically designed to investigate how RNAi works but also from observations made from its use as a technology in functional genomic studies. Greater mechanistic understanding has also contributed to subsequent improvements in the use of RNAi as a technology, most notably in human cell lines with profound implications for novel approaches to gene therapy (Billy et al., 2001; Elbashir et al., 2001a; Zamore, 2001).

5.2. The genetics and biochemistry of RNAi

Fire and associates were the first to report RNAi when they demonstrated the potent and specific disruption of gene function with the injection of dsRNA into *C. elegans* (Fire et al., 1998). There results catalysed a massive effort to understand the mechanisms underlying RNAi, PTGS observed in plants (Mourrain et al., 2000) and the 'quelling' phenomenon observed in fungi (Fagard et al., 2000). A complementary approach of biochemical and genetics approaches has rapidly unravelled the mysteries surrounding RNAi so that the current best model of how RNAi works is constantly changing (Lipardi et al., 2001; Nishikura, 2001). In addition to its roles in genome surveillance the genes involved in RNAi have now been shown to function in endogenous developmental processes (Ketting and Plasterk, 2000). Thus it appears

that RNAi has been co-opted from its original protective role to be involved in important regulatory functions.

Genetic studies have identified a number of genes involved in RNAi and subsequent biochemical studies have begun to confirmed their role and function in this process as well as describe the molecular events involved. These different studies have led to the most current model (at the time of writing) presented in Fig 5.1 and described below in the context of the most influential studies that have led to its proposal.

The first requirement for RNAi is presence of dsRNA within cells, whether by uptake, the production of dsRNA by viruses or transposable elements, or the endogenous formation of regulatory dsRNA duplexes (Fire, 1999; Fire et al., 1998; Kennerdell and Carthew, 2000; Kuwabara and Coulson, 2000a). This first part of the process has become known as the 'trigger' for RNAi (Fire, 1999). Any interference with the structure of the dsRNA trigger, such as sequence divergence or secondary chemical modification substantially lowers the potency of RNAi or abolishes it altogether (Parrish et al., 2000). The next part in the process of RNAi is the production of small fragments of dsRNA originating from the trigger molecules introduced into the cell cytoplasm (Elbashir et al., 2001b). This is mediated by the RNaseIII-like factor Dicer required for the cleavage of the trigger dsRNA into small dsRNA molecules of between 21 and 23 bp in length small interfering RNAs (siRNAs) (Elbashir et al., 2001b; Knight and Bass, 2001).

Genetic screens for RNAi resistant mutants and subsequent detailed functional analysis in *C. elegans* have identified the *rde-4* gene as a second factor essential for the degradation of trigger dsRNA into siRNAs (Parrish and Fire, 2001; Tabara et al., 1999). The requirement for siRNAs and an RnaseIII-like activity was originally demonstrated using a *D. melanogaster* extract system shown to be active for RNAi (Bernstein et al., 2001; Hammond et al., 2000). In addition this system was used to show that production of siRNAs from the dsRNA trigger leads to homology-dependent degradation of target mRNAs at 21- 23 nt intervals (Elbashir et al., 2001b). These observations led to an initial model for RNAi where dsRNA is processed to siRNAs that direct the cleavage of mRNA through sequence

complementarity. Subsequent to this biochemical analysis of RNAi the *C. elegans* homolog of DICER, *dcr-1*, has been shown to be required for RNAi in-vivo (Knight and Bass, 2001). In addition this gene was shown to be essential for normal development of the germline suggesting that dsRNA cleavage (and possibly RNAi) is a prerequisite for normal developmental processes (Knight and Bass, 2001).

Further biochemical studies have begun to identify the proteins involved in the targeted degradation of mRNA. Destruction of mRNAs homologous to the silencing trigger is mediated by the RNA-induced silencing complex (RISC) (Sharp, 2001). RISC is a sequence specific multi-component nuclease of approximately 500 kD that co-fractionates with 21-23 nt RNAs that are derived directly from the silencing trigger (Hammond et al., 2001). One component of RISC that has been identified is a member of the Argonaute family of proteins (Hammond et al., 2001). Genetic studies in *C. elegans*, *Neurospora* and *Arabadopsis* have identified members of this gene family that are essential for dsRNA mediated gene silencing (Fagard et al., 2000). Additional possible components of the RISC complex that would be expected from its biochemical activities would be both helicases and exonuclease activities. Concurrent with this the *mut-7* gene of *C. elegans*, a homolog of genes known to have both 3'-5' exonuclease and helicase activities, has been shown to be necessary for RNAi (Grishok et al., 2000; Ketting et al., 1999).

Most recently the role of RNA-directed RNA polymerases (RdRP) in the RNAi process has been investigated (Lipardi et al., 2001; Sijen et al., 2001). One intriguing observation in plants and *C. elegans* and more recently in *D. melanogaster* is that only very small amounts of trigger dsRNA are needed in a cell to induce RNAi (Cogoni and Macino, 1999; Fire et al., 1998; Kennerdell and Carthew, 2000). This is partly explained by the amplification implicit in the conversion of longer trigger RNAs into many siRNAs. However, this amplification is not sufficient to explain both the potency and self maintaining nature of RNAi observed in *C. elegans* (Fire et al., 1998).

This led to the proposal that an amplification step involving an RdRP might be an integral part of the RNAi process and genes with this activity have been identified by genetic screens (Cogoni and Macino, 1999; Sijen et al., 2001).

However, these genes have been shown not be essential for all aspects of RNAi, for example *ego-1* an RdRP from *C. elegans*, has been shown only to be required for RNAi in germline cells. One proposed role of RdRP activity was the amplification of small amounts of trigger dsRNA once it had entered cells. However, the demonstration that the production of asymmetric trigger dsRNA with chemical modifications of the antisense strand interrupted RNAi (whereas sense strand modifications did not) disagreed with this hypothesis (Parrish et al., 2000). If RdRP activity was indeed required to amplify initial small levels of dsRNA trigger interfering chemical modifications would be diluted out by amplification within the cell and would not preclude the RNAi effect. These observations added to the fact that no RdRP like genes have been found in the sequenced portions of the *Drosophila* or mammalian genomes put the role of RdRP activity seriously in doubt.

However genetic and biochemical studies investigating the somatically expressed RdRp gene *rrf-1* from *C. elegans* and biochemical studies in *D. melanogaster* have conclusively demonstrated the necessity of RdRP for RNAi (Lipardi et al., 2001; Nishikura, 2001; Sijen et al., 2001). Rather than working on the trigger dsRNA, RdRP activity appears to work by using siRNAs as primers to produce dsRNA duplexes with target mRNAs. Biochemical characterisation of siRNAs shows that they have a 3' hydroxyl group that would be necessary for this process (Sijen et al., 2001). RdRP activity is implied by the observation of secondary siRNA molecules that are complementary to the target mRNA in regions not contained in the original trigger dsRNA (Sijen et al., 2001). These secondary siRNAs are presumably generated by RNase III/DICER like activity on the mRNA/RdRP extension product duplex. In addition to amplification of specific target degredation these secondary siRNA molecules can lead to 'transitive' RNAi whereby they degrade other mRNA to which they are complementary and not targeted by the original trigger (Sijen et al., 2001).

This current working model of RNAi will no doubt be subsequently modified and outdated before long. Nonetheless the elucidation of the mechanisms behind RNAi add fuel to the hope that it can be used as a tool to study gene function directly in organisms where it was not previously possible. In particular the presence of a

powerful amplification mechanism during RNAi suggests that larger metazoans may be susceptible to this technique (and indeed are).

5.3. RNAi related mechanisms are required for normal functioning of essential developmental mechanisms.

Two observations led to the intriguing possibility that RNAi related mechanisms might be essential to normal developmental processes. Work in *C. elegans* had identified two small temporal RNAs (stRNAS, *let-7* and *lin-4*) that when mutated led to heterochronic phenotypes in *C. elegans*, effecting stage-specific development characterised by cell divisions that mimicked the pattern of division seen in earlier lifecycle stages (Reinhart et al., 2000). It was later shown that these stRNAs induced developmental progresssion by negatively regulating the expression of proteins by binding mRNAs with 3' untranslated regions (3' UTRs) of complementary sequence (Reinhart et al., 2000). The stRNAs bear remarkable similarity, in size at least, to the siRNAs responsible for RNAi (Elbashir et al., 2001a; Elbashir et al., 2001b). A second observation was that mutations in the *C. elegans* DICER gene, *dcr-1*, resulted developmental phenotypes, suggesting that RNAi related mechanisms may be involved in normal development (Knight and Bass, 2001).

Subsequently it has been demonstrated that the maturation of *lin-4* and *let-7* require not only *dcr-1*, but also members of the Argonaute protein family, *alg-1* and *alg-2* (Argonaute like genes), previously implicated in RNAi (Grishok et al., 2001). This clearly showed that common machinery guides both RNAi and some endogenous gene regulatory processes.

These observations were however only the tip of the iceberg as far as discerning the importance of this link. Techniques used to originally clone siRNAs from *D. melanogaster* identified a number of small RNAs of similar size that were complementary to sequences from the *D. melanogaster* genome (Elbashir et al., 2001b). These RNAs looked like the stRNAs already discovered in *C. elegans* by mutational analysis. This led to a concerted effort to clone stRNAs from the genomes of *C. elegans, D. melanogaster* and from humans (using the Hela cell line)(Lau et al.,

2001; Lee and Ambros, 2001). This has led to the identification of over 80 different 21-24 nt RNAs collectively called micro RNAs (miRNAs), in some cases conserved over three different phyla and with developmental specific expression profiles. Thus it appears that RNAi related mechanisms have been co-opted, during animal evolution at least, to form the basis of developmental gene regulation at the mRNA level.

Figure 5.1 A current model of RNAi

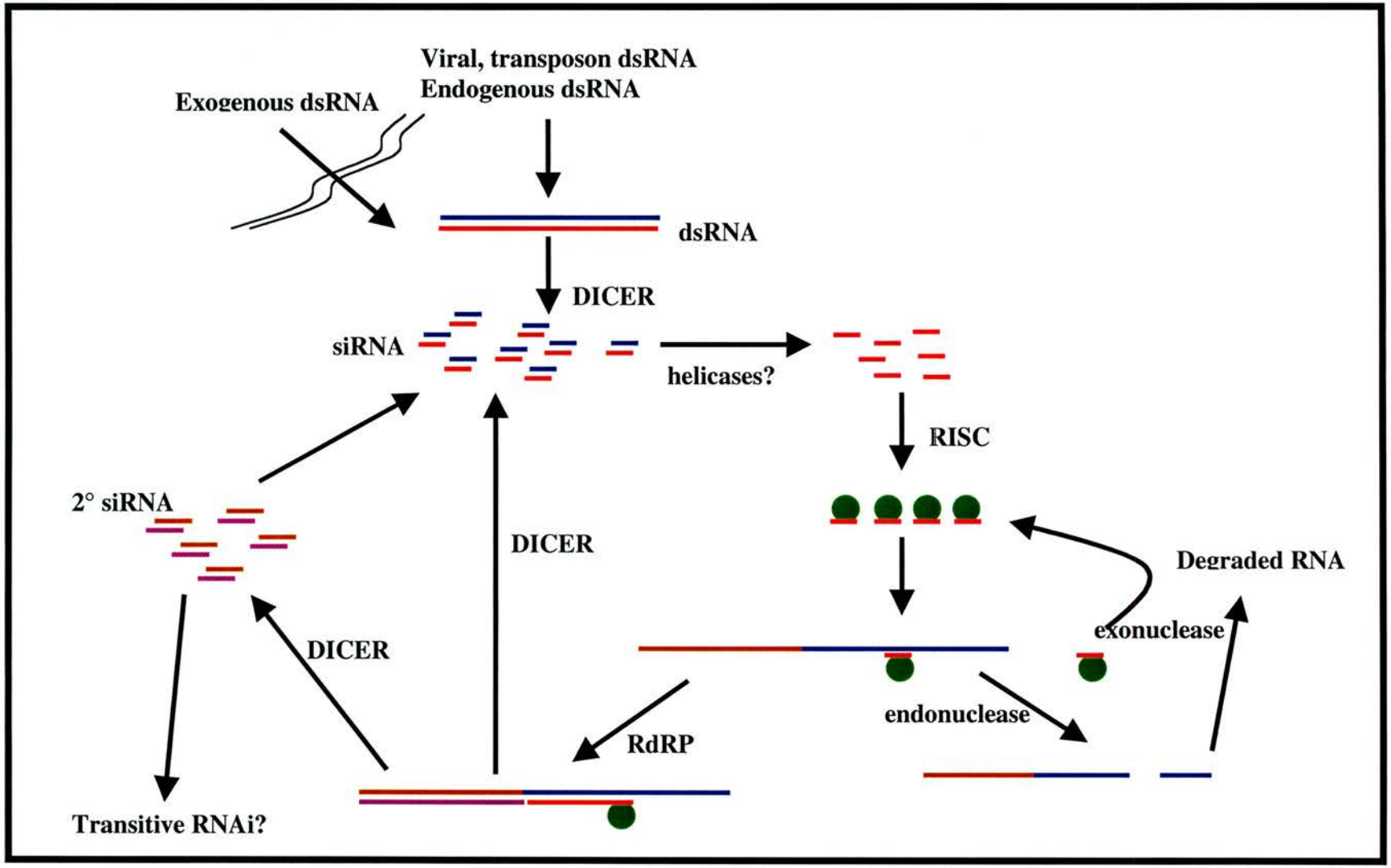


Figure 5.1 A current model of RNAi

A current model of RNAi based a synthesis of models proposed by Lipardi et al 2001 and Sijen et al 2001. The model is described in the text of the previous section. This models includes a mechanism that can explain the potency of RNAi that is observed even with introduction of only minimal amounts of 'trigger' dsRNA. The activity of RdRP is able to amplify the RNAi effect in proportion with the amount of target mRNA present in the cell. The new model also predicts that the effect of transitive RNAi, that is the targeting of a second mRNA with homology to the original target outside of the homology with the trigger dsRNA, may be observed. Transitive RNAi has been observed within the context of the experimental system used by ref but its importance in normal RNAi is unclear.

It appears that RNAi originated as genome surveillance mechanism to protect against the deleterious effects of viruses and transposons. In addition RNAi related mechanisms are essential for normal developmental processes as they are required for the production of miRNAs. These miRNAs appear to regulate developmental processes at the mRNA level.

5.4. The application of RNAi to non-genetic organisms; parasitic helminths

5.4.1. General considerations for the development of RNAi in non-genetic organisms

RNAi has now been applied as a technology to study gene function in a number of organisms considered previously to be genetically intractable (Brown et al., 1999; Schoppmeier and Damen, 2001; Shi et al., 2000). The only pre-requisite requirement for RNAi is partial coding sequence from the gene of interest. With the expansion of genome projects and the generation of sequence data from species of medical and economic importance this requirement is often already fulfilled. From this starting point whether RNAi can be applied is dependent on two major factors.

Firstly, RNAi will only work if the genes involved in dsRNA directed RNA degredation are conserved within the species of interest. We would perhaps expect this to be almost always the case for such an apparently conserved function, with only rare exceptions to the rule within the Eukaryotes. However the study of gene function by RNAi on a genome wide scale has so far only been practically possible in *C. elegans*. This has been facilitated not only by the existence of a well-defined genetic system, the vast amount of molecular genetic techniques and complete genome sequence, but also by the specifics of RNAi technology in this species. The study of gene function in *C. elegans* by RNAi is aided by its heritability; the F1 progeny of a treated animal inherit the RNAi effect (Fire et al., 1998). In fact RNAi phenotypes have mainly been screened for in the F1 generation, although the development of high-throughput RNAi by "soaking" protocols has made the study of post-embryonic functions of pleiotropic genes entirely feasible (Maeda et al., 2001). The heritability of RNAi is not just due to persistence of dsRNA species in the F1 progeny but has distinct genetic requirements (Grishok et al., 2000). It also appears that there is a "spreading" mechanism for RNAi in *C. elegans* such that exposure of cells in the gut to small amounts or dsRNA by microinjection results in RNAi in the germline and gonad, and presumably the rest of the organism (with the exception of neural tissues that may be refractory to spreading) (Fire, 1999). It appears that this spreading phenomenon is not just due to passive processes but has a genetic basis as

wild isolates of *C. elegans* have been found that while susceptible to RNAi do not transfer the effect to their progeny if injected in the gut as opposed to the gonad (Femke Simmer, personal communication). It is not yet at all clear whether the amplification phenomenon observed during spreading in *C. elegans* occurs in any other multi-cellular animals where RNAi has been used. A similar phenomenon exists in plants allowing systemic resistance through RNAi related mechanisms to viruses that have only previously infected, for example, in one leaf (Palauqui et al., 1997; Voinnet et al., 1998). Alternatively, resistance to a virus with some regions of high sequence homology to a previous virus has also been observed (Voinnet, 2001). This phenomenon can be explained by the current model of RNAi presented in Fig 5.1.

In *C. elegans* therefore it is possible to phenocopy null mutations in genes at most stages in the lifecycle. In other systems RNAi has not been developed to the same potency. For example in mammalian cell lines transcript levels of targeted genes are substantially reduced by the introduction of siRNAs, but not abolished (Elbashir et al., 2001a). Developing RNAi approaches for larger multicellular organisms may be a problem if the "spreading" phenomenon observed in *C. elegans* and plants is not universal. As mentioned above this amplification step is probably partially dependent on a special class of RNA directed RNA polymerases (RdRp) amplifying the effect of the original trigger, and for example, there is no homolog in the sequenced portion of the *Drosophila melanogaster* genome. Although an activity that would be predicted to be due to an RdRp has been identified in *D. melanogaster* (Lipardi et al., 2001). In addition to this it has been hypothesised that the RNAi effect may spread across cells, effecting cells that have not been primarily exposed to a trigger.

Secondly, whether RNAi will work is also a question of accessibility and delivery method. The correct life stages of an organism must be accessible to the delivery of dsRNA. For free living species or parasitic species that can be simply cultured *invitro* accessibility is not a major problem. In these cases a delivery system must be designed to deliver dsRNA in effective amounts. In *C. elegans* four different delivery mechanisms have been used. Microinjection of dsRNA (Fire et al., 1998), soaking in dsRNA (Maeda et al., 2001; Tabara et al., 1998), feeding with bacteria

expressing dsRNA (Timmons et al., 2001; Timmons and Fire, 1998) and transgenesis with constructs designed to produce dsRNA, are effective. Other systems have also employed transgenesis, microinjection and transfection techniques. In trypanosomes electroporation of dsRNA into cells has provided yet another means of delivery (Shi et al., 2000). It is likely that with most species of interest one of these methods can be adapted for use to deliver dsRNA into cells. A greater problem will be whether enough cells or the right lifecycle stages can be affected.

It is not yet clear how optimistic we can be about the future applications of RNAi as a generic technique in analysing gene function (Kuwabara and Coulson, 2000b). There are two areas of research of particular interest to me where the application of RNAi to previously non-genetic organisms may prove to be the catalyst for rapid advances. Firstly, the study of the evolution of developmental mechanisms (and the understanding of the evolution of gene function generally) will benefit greatly from the study of null phenotypes of conserved genes in a broader spectrum of taxa than currently available. Secondly, RNAi may prove to be the basis for studying gene function in (human) parasitic species at a level previously not possible. Many multi-cellular helminth parasites are not amenable to any genetic study because of the complications of maintaining experimental lifecycles in the laboratory. Thus far attempts at transgenesis have either failed or not proved successful enough to be of any practical use. Studies of gene function have therefore been limited to the expression and analysis of proteins of putative importance to the parasitic mode (Gomez-Escobar et al., 1998; Zang et al., 1999). RNAi may allow for the first time the study of the role of these genes in the context of the currently established laboratory models of infection, allowing the identification of essential genes and putative drug targets.

5.4.2. The potential application of RNAi to parasitic helminths

Nematode parasites of humans, animals and plants cause premature death, chronic sickness, and loss of productive labour, widespread malnutrition and impose a multibillion-dollar load on developed and developing countries. Estimates of the number of major human helminth infections are currently running at over 4 billion totalled together, including around 500 million attributed to non-nematodes

(Crompton 1999). Some of the nematodes responsible for these infections are closely related to *C. elegans*, while other species are more distant. *C. elegans* has proved its worth as a model organism and will continue to be used extensively. Surprisingly, in my opinion, it has not been used as a model for parasitic nematode biology to any great extent, probably because of the success of its use in answering universal questions (Aboobaker and Blaxter, 2000).

A robust molecular evolutionary framework exists for the phylum Nematoda and from this it is apparent that some parasitic nematodes are much more closely related to *C. elegans* than others (Fig 5.2). We might expect more chance of successfully employing RNAi in these parasitic taxa than those distantly related to *C. elegans*. However, unpublished data from work on free-living nematodes closely related to *C. elegans* is not encouraging. For example attempts at RNAi in *Pristionchus pacificus*, a member of the closely related Diplogasterida being developed as a satellite model to *C. elegans*, have so far failed (Ralf Sommer, personal communication). But more encouragingly initial attempts at RNAi in the vertebrate gut parasite *Nippostrongylus braziliensis*, also relatively closely related to *C. elegans*, have been very encouraging (Murray Selkirk, personal communication). Considering the conservation of RNAi across Eukaryota it is likely that all species are capable of RNAi except where required genes have been lost, perhaps if no selective pressure to maintain an active RNAi process is present. This might be the case for taxa where there are no infective or deleterious viruses or active transposable elements. However the growing body of evidence that many of the genes involved in RNAi also control the expression of endogenous genes by similar mechanisms and have other essential functions suggests this is unlikely.

The application of RNAi to parasitic nematodes could be an incredibly powerful tool for studying gene function. Most significantly if RNAi can be combined with sensitive assays it should be possible to identify novel proteins required for parasite survival either within the host or vector species if there is one. The identification of such genes, some of which are likely to lack close homologs in the host, would be ideal candidates for novel drug targets. This should also hold true for other parasitic species closely related to the species under direct study.

Fig 5.2 The phlyogenetic structure of the Nematoda

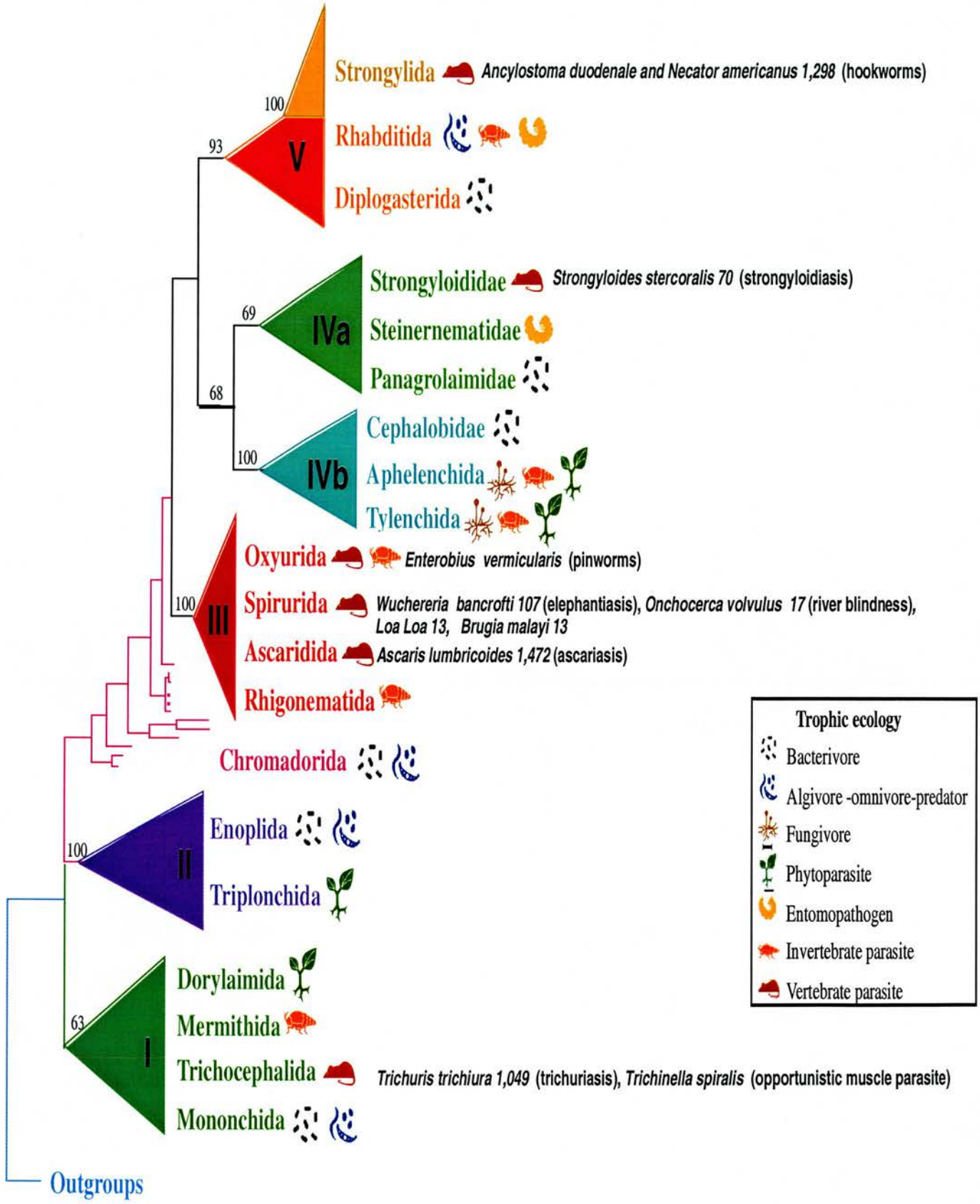


Fig 5.2 The phlyogenetic structure of the Nematoda

A phylogenetic analysis of full-length 18s rDNA sequences from the phylum Nematoda showing the relationships between the major clades. Numbers at the branches of the tree indicate percentage bootstrap support for these branches (after Blaxter et al 1998). There are important human pathogens in 4 of the 5 major groups (I-V). The numbers after the species name (where present) indicate the number of estimated current infections in numbers of millions. *Enterobius* (pinworm) is a cosmopolitan infection of developed and developing nations and it is likely to infect most people. The parasitic *Strongylids* are the parasitic group most closely related to *C. elegans*. RNAi has been successfully used in the rat gut parasite *Nippostrongylus braziliensis* (Murray Selkirk, personal communication) a member of the *Strongylids*. (Adapted from Aboobaker and Blaxter, 2000).

5.5. The filarial parasite *Brugia malayi* as a candidate for an RNAi approach to studying gene function and screening for drug targets

5.5.1. Brugia malayi, the filarial genome project and the need for a novel approaches to study gene function

Lymphatic filariasis is a chronic, debilitating disease resulting in high morbidity among the 120 million people affected by it worldwide. *Wuchereria bancrofti*, *Brugia malayi* and *Brugia timori* are the species of the filarial parasitic nematodes that infect humans (Maizels et al., 2001). Of these *Brugia malayi* can be cultured through its lifecycle through both mosquito vector (*Aedes aegypti*) and mammalian host (using the Mongolian jird *Meriones unguiculatus*) and has thus been the subject of most investigation. *Brugia malayi* is the subject of an advanced genomics effort by the Filarial Genome Project (FGP), co-ordinated by the World Health Organisation, that has generated over 23,000 expressed sequence tags (ESTs) from cDNA libraries spanning the lifecycle (Fig 6.3), representing approximately 8,000 different open reading frames (ORFs) (John Parkinson, personal communication). This is predicted to represent over one third of the genes present in the 100Mb genome by analogy with *C. elegans*. The FGP has greatly aided more classical approaches to understanding the biology of the parasite, clearly identifying novel genes that, either through their high expression levels or sequence similarity can be predicted to have an important role in parasite biology (Gomez-Escobar et al., 1998; Zang et al., 1999).

Nevertheless this wealth of information has only aided the initial identification of candidate drug targets or vaccine candidates (Gregory et al., 2000; Manoury et al., 2001; Murray et al., 2001). Functional studies of these genes continue to be based upon a biochemical gene-by-gene approach. As the functions of proteins are studied either in biochemical assays *invitro* or in a heterologous *invivo* system in the context of an effect upon or response by a mammalian host, it is often impossible to be sure of true function. For some *Brugia* genes of interest it is

possible to identify *C. elegans* homologs and study in the tractable model organism may provide insight into parasite gene function. However the fact remains that *C. elegans* is not a parasite and is only distantly related to *B. malayi* so any insight into function must be treated with caution.

In order to both increase the speed at which novel genes are studied and also understand more clearly what their roles are, new approaches to studying gene function are required. As *B. malayi* remains an intractable genetic system despite its position as a model filarial parasite, RNAi potentially represents a way to study gene function. If the mechanism of RNAi is conserved it should be possible to exploit it to study the functions of some genes directly either during in-vitro culture or by performing RNAi and reintroducing parasites back into the lifecycle. However in order to exploit RNAi in this way a number of technical difficulties associated with working with a parasite will have to be overcome and suitable primary assays developed to test that any affects of knocking out gene function can be measured.

5.5.2. A rationale for RNAi in Brugia malayi

There are four major problems that need addressing in order to test whether RNAi works in *B. malayi*. These inter-related problems are a suitable delivery method, choice of lifecycle stage with which to work, choice of initial target genes and initial molecular and phenotypic assays to assess any affects.

A consideration of the *Brugia malayi* lifecycle and how it is maintained in the laboratory suggest that three lifecycle stages are available for experimentation. On sacrifice of patent jird hosts both adults and the blood stage microfilaria (Mf) are available and, on preparation of mosquito vectors around 2 weeks after they have been blood fed microfilaria, infective L3s are available. Of these, infective L3s are the scarcest resource for maintenance of a successful lifecycle and therefore have limited availability. The intervening stages of the lifecycle can be obtained from both vector and jird hosts by correctly timing preparation and sacrifice. In the experiments here only adults and Mf have been investigated. These are the easiest stages to obtain

from the lifecycle and offer a number of advantages in terms of downstream analysis of experimental results (see below).

Of the delivery methods used for *C. elegans* and those in other systems only one is suitable for *B. malayi*. Microinjection of *B. malayi* is difficult because of high hydrostatic pressure resulting in most worms exploding (Professor Alan Scott, personal communication). As adult female *B. malayi* are so much bigger then *C. elegans* it may be that the small absolute amounts of dsRNA delivered would not have any affect. *B. malayi* does not appear to ingest orally in culture and certainly does not eat *E. coli*. The remaining option by analogy with *C. elegans* is to perform RNAi by soaking in high concentrations of dsRNA. The concentrations used for *C. elegans* are in the region of 3-5mg dsRNA per ml in a volume of 10µl (Tabara et al., 1998). As *C. elegans* is so small relative to *B. malayi* (1 mm long compared to approximately 30-90 mm long) much larger amounts of dsRNA would be needed to allow soaking of *B. malayi* in suitable volumes. In addition *B. malayi* has exacting requirements in culture and confinement to small volumes for extended periods of time without replenishment of nutrients is likely to be detrimental. For RNAi by soaking to work these technical difficulties will have to be overcome.

To assay whether RNAi has worked it is necessary to assess endogenous levels of target transcript mRNA. The most sensitive way to do this would be to compare the transcript levels of treated and untreated worms using optimised single worm RT-PCR reactions. If reductions in transcript level are observed protein levels can then be observed by Western blot analysis if suitable reagents are available. These assays should establish a timeframe for how long RNAi might take and how long its effects last for. Initial gene targets for RNAi should have easily predicted phenotypes on worms in culture to indicate that RNAi has been successful. The phenotypes that can be observed in culture are limited. Adults and Mf in culture are highly motile when healthy and this could be used if the phenotype of untreated worms in culture is consistent enough to allow comparison. In culture fertilised embryos continue to develop in adult females and are released as Mf into the surrounding culture medium. RNAi of genes predicted to be involved in this process could result in abnormal Mf phenotypes or decreased Mf release. Perhaps the most

obvious phenotype for establishing RNAi as a technique would be to induce worm death in culture by performing RNAi of an essential gene.

The rationale outlined above should be an efficient approach to establishing whether RNAi works in *B. malayi*, if simple effective assays for the function of some genes can be established in culture and how useful a technique RNAi could be for further studies of gene function in this parasite.

Fig 5.3 Outline of the lifecycle of the filarial parasite *B. malayi*

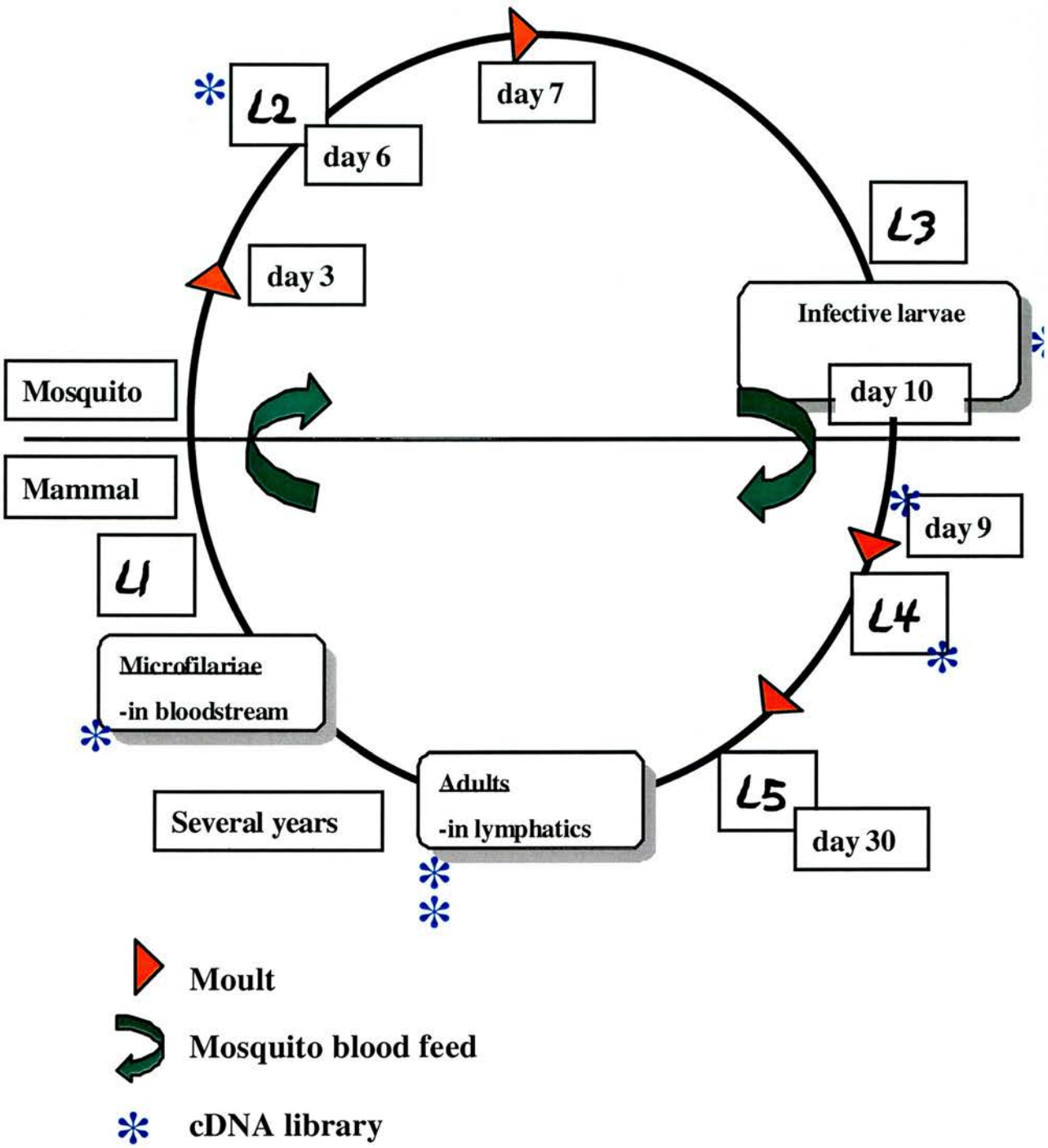


Fig 5.3 Outline of the lifecycle of the filarial parasite *Brugia malayi*

The lifecycle of the filarial parasite *Brugia malayi* makes classical and molecular genetic analysis impossible. Mature adult female parasites release millions of microfilaria into the host bloodstream and can survive for several years. Circulating microfilaria in the blood are taken up by the vector mosquito host during blood meals. Microfilaria migrate from the mosquito mouthparts to the flight muscles by travelling through the gut wall, where they continue to mature. After three days Mf moult to the L2 larval stage and then undergo a second moult after seven days to the L3 larval stage. Between days 9 to 12 L3 larvae competent to re-infect the mammalian host migrate to the mosquito mouthparts. Infective larvae are reintroduced to the mammalian host when infected mosquitoes blood feed. Infective larvae mature through two further moults and adult infection establishes in the lymphatics. Mature adult males are then able to fertilise mature adult females, which then begin to release microfilaria. The WHO coordinated filarial genome project has sequenced over 23,000 ESTs from the different cDNA libraries prepared from around the lifecycle of the parasite (Blaxter et al., 1999).

For the purposes of RNAi the life cycle stages most accessible in the laboratory are adult parasites and microfilaria. Also possible is the targeting of genes required for embryogenesis and microfilarial development *in-utero* (Rogers et al., 1976), Interfering with the function of these genes at the correct time should produce visible phenotypes in the Mf released by maturing adult females.

5.6. Materials and Methods for RNAi in *Brugia malayi*

The materials and methods described here are those that have been used or developed specifically for RNAi experiments in *B. malayi*. All other materials and methods mentioned in this chapter are described in detail in the materials and methods chapter of this thesis.

5.6.1. Standard culture of Brugia malayi

Adult female parasites were a kind gift from Professor Rick Maizels and were taken from the lifecycle currently maintained by Ms Yvonne Harcus at the Institute of Cell and Animal Population Biology. Parasites were removed form the peritoneal cavity of sacrificed jirds (*Meriones unguiculatus*) and washed once in standard culture medium (RPMI 1640 (Invitrogen Life Technologies,UK), 1% penicillin/streptomycin (10,000 units/ml Invitrogen Life Technologies,UK) 2% glucose) preheated to 37°C. Parasites were then transferred to standard culture media of a total volume of 50-150 ml depending on the total number of parasites (about 5ml per parasite). This medium was changed every 12 hours for parasites being used for subsequent qualitative or quantitative analysis.

5.6.2. Small volume culture of Brugia malayi

Adult female parasites were cultured in 12 well microtitre plates in groups of 2 or 3 in 2ml of standard culture medium and 1mM spermidine (Invitrogen Life Technologies,UK) at 37°C for 24 hours (hr). The medium was supplemented when required with a 50mM $NaHCO_3$ solution in drop-wise fashion until the medium turned red indicating a return to neutral pH. After 24 hr parasites were maintained in their pairs or groups and moved to normal culture conditions as described above for continued observation.

5.6.3. Culture of Brugia malayi in a dialysis system

The domed tops of 300µl PCR tubes (Advanced Bio, UK) were first removed with a sterile scalpel. A piece of dialysis tubing with a molecular weight cut off of 15 kDa (Spectrapore, USA) was cut to single thickness and placed between the lid and the tube to form a barrier between the tube contents and the exterior. Tubes were filled with approximately 330 µl of standard culture media, and groups of 6,7 or 8 healthy adult female parasites (assessed by high motility) were added with or without either FITC-labelled dsRNA or dsRNA. Tubes containing parasites were then placed in beakers containing 100 ml of standard culture medium preheated to 37°C.

During the next 24 hr period tubes were checked every two hours to assess worm health by motility and to check that dsRNA had not precipitated in the tubes to which it had been added. In any case the contents of all tubes was mixed every 4 hr by gently pipetteing the contents. The volume in the tube was then transferred to a fresh 1 ml tube and spun at 5000 rpm for two minutes in a bench top centrifuge to pellet released Mf. The supernatant media was then returned to the original tube and a fresh piece of dialysis tubing was placed over the top of the tube and the lid closed. During these manipulations (which take approximately 5 minutes for each tube) the parasites were transferred temporarily to beakers containing standard culturing medium and then put back into their original tubes containing the Mf free culture medium and dsRNA when present. Tubes were then returned to the original media containing beakers. In some experiments individual parasites were removed at required time points, rinsed briefly in PBS and frozen in Trizol reagent (Invitrogen Life Sciences, UK) for later preparation of nucleic acid and protein. After 24 hr the (remaining) parasites were placed in normal culture conditions for continued observation, accept in the cases when they were already dead.

5.6.3. Generation of templates for in-vitro transcription

Approximately 300 bp fragments of the genes under investigation were cloned into a vector with T7 and T3 promoter sites (TA Topo cloning kit, Invitrogen Lifesciences.

UK). Approximately 10 μg of DNA template was generated from these constructs by performing several 100 μl PCR reactions using T7 and T3 vector primers the conditions used for screening for positive colonies from cloning reactions (see Materials and Methods), gel purified and used as templates for *in vitro* transcription.

5.6.4. Generation of dsRNA and FITC-labelled dsRNA

FITC-labelled dsRNA was generated using FITC conjugated uracil (Invitrogen Life Sciences, UK) and performing separate T7 and T3 polymerase (Invitrogen Life Sciences,UK) in-vitro transcription reactions on 1ug of template DNA according to the manufacturers protocol. Each reaction was treated with 5 units of RNAse free DNAse I (Invitrogen Life Sciences UK) to remove template DNA. The reactions were then pooled, heated to 60°C for 5 minutes and allowed to cool at room temperature to allow annealing of complementary strands. A small aliquot was then run on a standard agarose gel to check for integrity of dsRNA. The FITC-dsRNA was then dialysed over RNAse free water to remove free nucleotides and quantified using a spectrophotometer (Genequant, Abgene, UK).

Generation of large quantities of dsRNA using larger scale in-vitro transcription reactions were performed using the Ambion Mega Script kit (Ambion, USA) according to the manufacturers protocol. Concentrations of PCR generated template producing the highest yield of RNA were found to be different for each template and were discovered empirically by performing smaller scale reactions, and quantifying the amount of RNA produced using a spectrophotometer (Genequant, Abgene, UK). These reactions were then scaled up such that each T7 or T3 polymerase reaction produced in the region of 600 μg of RNA. T7 and T3 reactions from the same template were then pooled after purification according to the manufacturers protocol, heated to 60 °C for 10 minutes and left to cool at room temperature to allow complementary strands to anneal after template DNA had been removed by DNAse I treatment (10 units at 37 °C for 30 minutes). Integrity of the dsRNA was checked by running on a standard agarose gel and then stored precipitated under 100% ethanol in dome capped PCR tubes at –20 °C for up to 2

days before use in RNAi experiments at which point it was re-suspended in standard culturing medium with 1 mM spermidine prior to parasites being added.

5.6.5. Quantification of microfilarial release in culture

Quantification of Mf release in culture was performed with either individual parasites or parasites pooled into groups. Individual worms were cultured for 1-3 hr in 10 ml of standard culture medium in 15 ml Falcon tubes at 37 °C. The 10 ml of culture medium was then centrifuged in 15 ml Falcon tubes at 3000 rpm after the parasite was removed to gently pellet all the released Mf. The top 9.8 ml of media was removed and the Mf re-suspended in the remaining 200 µl. Mf counts were then performed on five 20 µl aliquots of this remaining 200 µl under a binocular light microscope at 100 x magnification. Counts were adjusted to represent the Mf release per hour depending on the original collection time.

Mf release from groups of parasites was performed in the same way except that parasites were cultured in 50 ml falcon tubes in 40 ml of standard culture media. An alternative method for Mf counting was also used when Mf phenotypes were visualised concurrently and is described below. This was used in addition to the method described here and not as an alternative.

5.6.6. Visualisation of microfilarial phenotypes released in culture

Mf were collected for a set period of time, usually three hours as described above. After centrifugation and re-suspension in culture media 50 µl was placed on a glass microscope slide and spread gently using a 200 µl pipette tip. Slides were then left to air dry for 20 minutes and then fixed in 100% methanol for 5 minutes followed by rinsing in ddH_20. After slides had been allowed to air dry for a further 10 minutes they were placed in Giemsa stain (Sigma, UK) diluted 1:10 in ddH_20 for 20 minutes. Slides were washed again for 5 minutes in ddH_20. After being allowed to dry at 4 °C overnight slides were overlaid with large cover slips.

Mf were viewed under an inverted light microscope, the number of Mf on the slide were counted and any unusual phenotypes scored and recorded. Pictures of microfilaria were taken using a colour digital camera and the Openlab 3.0 software. Images were viewed and stored on a G3 PowerMac computer.

5.6.7. Single worm RT-PCR

Single adult female parasitic worms were removed from culture and placed in Trizol reagent (Invitrogen Life Sciences, UK) and total RNA prepared according to the manufacturer's protocol (also described in materials and methods). Remaining material was retained to allow later protein extraction. The extracted total RNA was DNAse I treated and used in a single oligo-dT primed reverse transcriptase reaction using the Stratagene High Fidelity RT-PCR Kit (Stratagene, UK). The resulting 15 µl cDNA reaction was dialysed over ddH_20 and re-concentrated to 12 µl using a DNA vacuum centrifuge. The expression levels of genes of interest were than assessed by 20 µl PCR reactions using 1, 2 and 3 µl of cDNA template. PCR reactions were optimised for each gene, such that 3µl of PCR product run on a 1.2% agarose gel gave strong bands when 3 µl of cDNA was used as template and progressively weaker bands when 2 µl or 1 µl was used (see Table 5.1 for PCR conditions for each gene, see Appendix 2 for gene and primer sequences).

5.6.8. Fluorescent microscopy of whole mount parasites

Single adult female parasites and both sheathed and ex-sheathed Mf soaked in FITC-dsRNA for a 24 hr period in the dialysis system described above were removed from culture medium and briefly rinsed in PBS. Parasites were then fixed for 5 minutes in 4.2% formalin and then placed on glass sides pre-treated with poly-lysine and allowed to air dry for 10 minutes. Slides were then overlaid with large cover slips and sealed with clear nail varnish.

Parasites were viewed under a fluorescent microscope first under an alternative fluorescent wavelength and then under the wavelength for FITC and images captured using a digital camera and the Openlab 3.0 software for fluorescent microscopy. The Openlab 3.0 software was used to subtract background levels of fluorescence observed under all wavelengths from captured digital images.

5.6.9. Ex-sheathment of microfilaria in culture

Mf were pooled and resuspended in 2 ml Hank's Buffered Salt Solution (HBSS, Invitrogen Life Technologies,UK) and allowed to stand for 5 minutes. To remove the proteinaceous sheath 2 ml of 2 mg/ml of Pronase solution (Sigma Protease type XIV from *S. griseus* suspended in HBSS) were added to a final concentration of 1 mg/ml. Mf were left to ex-sheath at room temperature in a 15 ml centrifuge tube for 10 minutes with gentle rolling. After 10 minutes 400 µl of 10% Fetal Calf Serum (FCS, Invitrogen Life Technologies UK) was added. Mf were then centrifuged at 3000 rpm for 5 minutes and washed twice in 9 ml HBSS and 1 ml of 10% FCS, centrifuging for 5 minutes after each wash. Mf were washed for a final time in 10 ml HBSS. A sample of Mf were transferred to glass slides and ex-sheathment checked by light microscopy. Around 90-95% of Mf were found to be successfully ex-sheathed by this treatment.

Table 5.1 Optimised PCR conditions for assessment of single worms transcript levels

Gene name	PCR cycle
shp-1	**94°C 5 min x 1** **(94°C 15 sec, 55°C 15 sec, 72°C 30 sec) x 29** **72°C 3 min x 1**
Bm-tub-1	**94°C 5 min x 1** **(94°C 15 sec, 58°C 15 sec, 72°C 30 sec) x 30** **72°C 3 min x 1**
Bm-ama-1	**94°C 5 min x 1** **(94°C 15 sec, 59°C 15 sec, 72°C 30 sec) x 31** **72°C 3 min x 1**

Table 5.1 Optimised PCR conditions for assessment of single worms transcript levels

PCR conditions were optimised using first strand cDNA prepared from single adult female *B. malayi*. PCR conditions were optimised to allow the amplification of each gene such that the use of 1 µl, 2 µl and 3 µl of cDNA respectively resulted in an increase in the amount of PCR product easily detectable by agarose gel electrphoresis.

5.7. Demonstration of uptake of FITC-labelled dsRNA by adult female *Brugia malayi* in culture

In order to begin to develop a method to perform RNAi by soaking in *Brugia malayi* it was first necessary to demonstrate that dsRNA was taken up by parasites in culture. At the time these initial experiments were done it had been shown that pieces of dsRNA smaller than 200 bp did not always produce strong affects in *C. elegans*, although this has subsequently been proven not to be the case. For this reason molecules of slightly greater than 300 bp (including vector sequence) of dsRNA have been used in the experiments described here as it was thought that smaller dsRNA molecules might have a greater chance of entering *B. malayi*.

FITC –labelled dsRNA, produced by in-vitro transcription corresponding to the *Bm-tub-1* gene was taken up by parasites in culture (Fig 5.4). Spermidine was used in the culture medium as this protein has been shown to enhance RNAi by soaking in *C. elegans*, although the mechanism for this is not known. Levels of fluorescence were the same for different constructs and did not increase for periods of soaking beyond 18 hours.

These data clearly indicate that *B. malayi* will take up dsRNA in culture. Whether this is a passive or active process is not known. Staining appeared to be strongest at the mouth and vulval opening in the anterior of the worm and throughout the worm gut. This is not surprising as these openings are likely to present the least resistance to the entry of macromolecules and perhaps the only route of entry.

The finding that dsRNA of 300 bp in size can gain access to worms suggests that performing RNAi by soaking on *Brugia malayi* may be possible on adult female parasites. Similar experiments were also performed on Mf, both those released in culture and those collected directly from jird hosts. Mf failed to show any internal fluorescence when soaked in FITC-dsRNA. To test whether this was due to the presence of the protective sheath that surrounds *B. malayi* Mf the experiment was repeated with Mf that had been ex-sheathed. However, it was found that ex-sheathed Mf also failed to take up dsRNA in culture. Previous experiments have shown that proteins of up to 45 kDa in mass can penetrate the sheath of Mf and cuticle of adult

parasites (Devaney, 1985). However, many researchers consider Mf to represent an arrested developmental state suited to survival in host blood and that large macromolecules are not taken up either through the sheath or gut (Bill Gregory, personal communication). Feeding is assumed to occur by the uptake of smaller molecular mass nutrients conveniently provided by the host.

Fig 5.4 Demonstration of dsRNA uptake by adult female *Brugia malayi* parasites

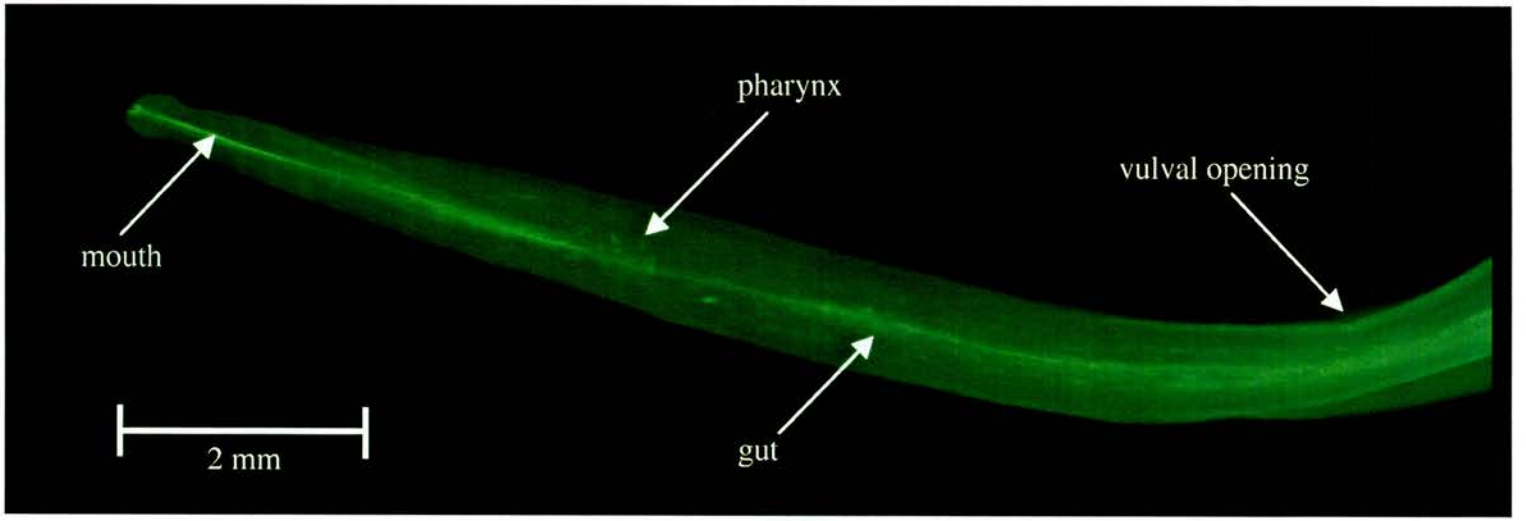


Fig 5.4 Demonstration of dsRNA uptake by adult female *Brugia malayi* parasites

Adult female *B. malayi* were soaked for 18h hrs in FITC-labelled dsRNA at a concentration of approximately 80μg/ml. Soaking was performed in the dialysis system described above to limit the amount of FITC-labelled dsRNA required. Fluorescence was visualised under a fluorescent microscope and digital software used to remove background fluorescence. Staining is strongest at the mouth and vulval opening at the anterior of the worm and throughout the worm gut. These data indicate that dsRNA is taken by adult parasites suggesting that soaking is a feasible method for delivering dsRNA for performing RNAi.

5.8. Development of a small volume dialysis culture system for an RNAi by soaking method in *Brugia malayi*

In order to perform RNAi by soaking at a useful scale it was first necessary to develop a culture method that allowed exposure of adult worms to high concentrations of dsRNA without affecting health and fecundity of the worms. Soaking experiments in *C. elegans* have typically used concentrations of 3-5 mg of dsRNA/ml with 4-5 L4 worms in a final volume of 10-15 µl for 24hrs, a total amount of 3-5 µg of dsRNA. By analogy soaking a single adult female *B. malayi* in 1.5 ml of culture media would require 3-5 mg of dsRNA to achieve the same concentrations. Brugia is usually cultured collectively in volumes in excess of 50 ml with either transfer to fresh media and/or addition of glucose and sodium bicarbonate to maintain the correct pH on a daily basis. Here two alternative methods were developed, one of which does not significantly affect parasite lifespan in culture or Mf release in culture. These measures of parasite health/fecundity appear to be robust phenotypes to observe under the culture conditions used here.

5.8.1. Microfilarial release by adult females in culture.

Mf release in culture by adult female *B. malayi* was investigated for two reasons. Firstly, to see if it could be used as a way to measure any deleterious affects, along with lifespan, of the small volume culture methods described below. Secondly to see if it was a robust enough phenotype to use to look at the effects of RNAi, with the downstream application of allowing the identification of genes required in one way or another for Mf release.

Mf releases of all the healthy individual adult females from three different jird hosts were counted for the 3 hours post-sacrifice. All three jirds used had in excess of 80 healthy adult individuals (assessed by high motility in culture and including both sexes) had been infected with infective L3 stage parasites between 8-14 weeks previously. Statistical analysis (Table 5.2) shows that Mf release from

these worms appears by eye to be normally distributed within populations taken from the same host (Fig 5.5 and Table 5.2).

These data indicate that Mf release in culture may be robust enough to use as a measure of any the deleterious affects of both small volume culture methods and RNAi as long as any comparative analysis between control and experimental conditions is done on parasites from the same jird host. Practically this limits its use as a phenotype to experiments that require up to perhaps 40 individual adult female parasites if they must all come from the same jird. This would allow Mf release from each worm to be calculated individually at the beginning of an experiment to assess the distribution of Mf release frequencies at the beginning of an experiment and to remove those individuals at the extremes of the distribution. The remaining individuals could then be pooled into groups with approximately equal average Mf release frequencies. Discounting individuals at the extremes of the distribution will make it easier to form groups that will make downstream statistical analysis more robust.

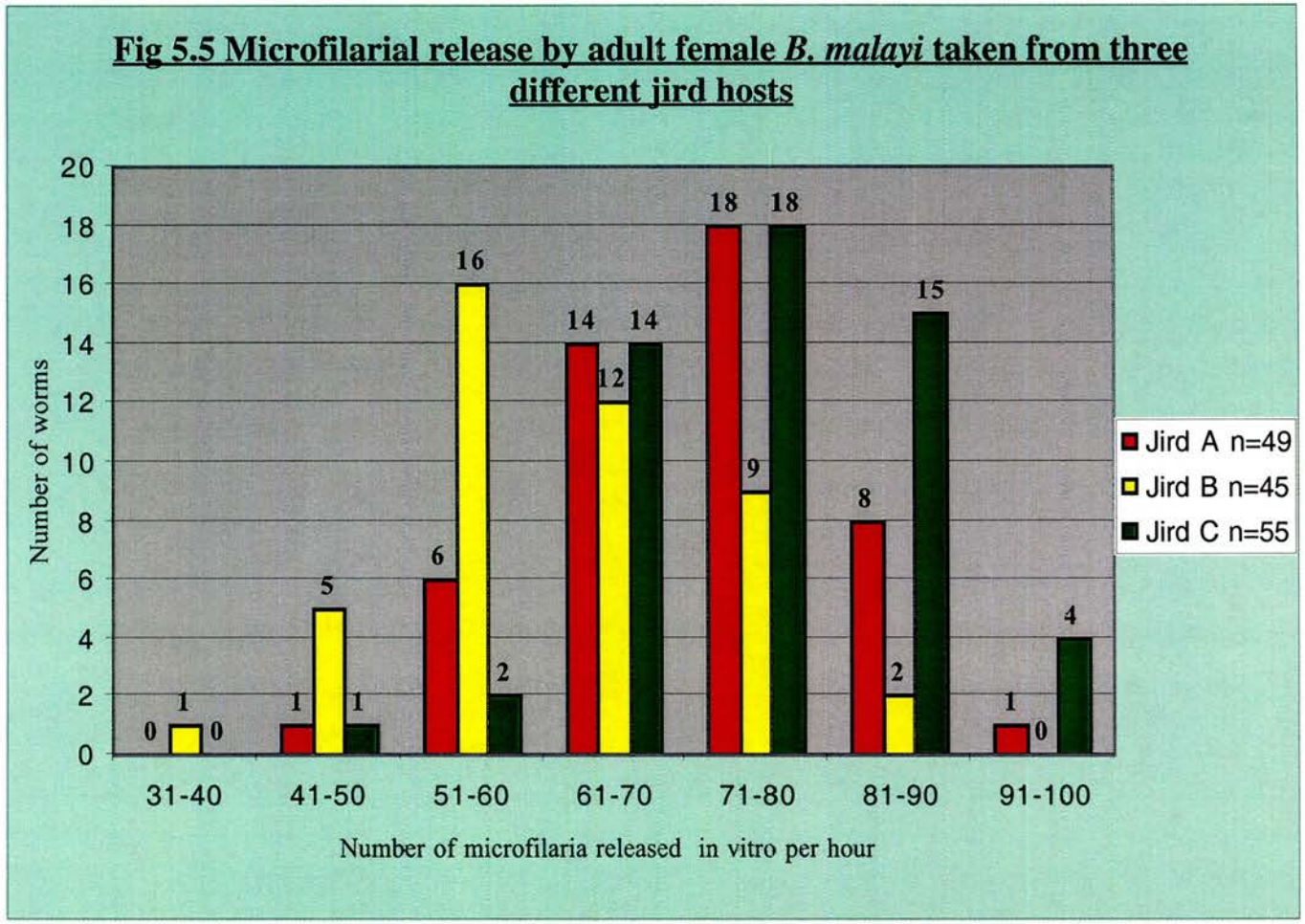


Fig 5.5 Microfilarial release by adult female *Brugia malayi* taken from three different jird hosts

Mf release from the adult female parasites from three different jird hosts was assessed in culture after sacrifice. Mf release was measured as described above and appears to be normally distributed within each jird host (see Table 5.2). Mf release is can be used as a measure of parasite health and fecundity in culture conditions as groups of individuals with approximately the same Mf release frequencies within and between groups can be formed.

Table 5.2 Microfilaria release is in discrete populations from different jird hosts

Jird	Mean Mf release per hour	Standard deviation of the mean
A	71	10.5
B	62	10.4
C	79	10.8

Table 5.2Microfilaria release is in discrete populations from different jird hosts.

The mean Mf release per hour was calculated along with the standard deviation of the mean for each jird host. The student T-test indicates that the Mf release frequencies from Jird A form a discrete population from those of Jirds B and C ($p<0.05$ in both cases) and that B and C are also from discrete populations ($p<0.01$). This indicates that while Mf release can only be used as readout for health and fecundity when comparing parasites originating from the same host.

5.8.2. Parasite lifespan and Mf release are not significantly affected by culture in a dialysis system for 24 hours post sacrifice

In order to expose adult female parasites to high concentrations of dsRNA in culture a considerable emphasis was placed on developing a robust culture system. Two different basic approaches were tried to solve this problem. One was to culture parasites in small absolute volumes of media for a fixed period. The other was to use a dialysis system which would allow culture in large volumes but maintain high concentrations of dsRNA. The advantage of the first system is that it reduces the amount of parasite manipulation during the experiment but results in culturing in less than ideal volumes of culture media. Dialysis culture systems circumvent this problem but require much greater parasite manipulation and are quite frankly very fiddly.

In order to assess the success of culturing methods two measures of parasite health were used. Firstly lifespan in culture was assessed by observing groups of worms in normal culture conditions. Worms were considered to be unhealthy if they displayed reduced motility and began to uncoil and straighten out in normal culture conditions; this phenotype normally indicates that worms will be dead within 72 hours. Worms were considered dead when they no longer moved and floated in culture media.

Culturing in small volumes of culture media for 24 hr using the method outlined above resulted in significantly reduced lifespan in culture (Fig 5.6). Based on this finding this method was not considered further as it obviously effects parasite health adversely and does not appear to be a robust method.

In contrast to this culture of parasites within the optimised dialysis system for 24 hr only resulted in a slight, but not significant reduction in lifespan when seven or fewer parasites were used (Fig 5.6). This slight reduction was consistently observed and could be due to the limits of diffusion set by the relatively small area of the dialysis membrane contact between the media in which the parasites are kept and the surrounding media. This problem may be circumvented by the design of a holding vessel allowing greater contact with the surrounding media. When eight or more

female parasites were used a significant reduction in lifespan was observed (Fig 5.6), suggesting that this number is too many to maintain parasite health in this system.

Mf release was also measured for parasites cultured in groups of seven for the days proceeding the 24 hr period in the dialysis system (Fig 5.7). The Mf release immediately post sacrifice of all the individual parasites in the control groups and those cultured in groups of seven within the dialysis system was measured. Adult parasites with Mf releases close to the mean of the whole population from each individual jird were pooled into groups so that each group had roughly the equivalent average Mf release frequencies. This figure was taken as the average Mf release frequency for each group for the first day post-sacrifice. After culturing for a further 24 hr (in either normal or dialysis conditions) Mf release for each group was measured again to give the Mf release average after 2 days post-sacrifice. This was repeated every 24 hours for each group until the first unhealthy worm was observed.

A slight decrease in Mf release was observed for those parasites maintained in the dialysis culture system for the first 24 hours post–sacrifice (Fig 5.7). This slight reduction most likely represents a small effect on worm health (also observed for lifespan measurements) due to culture in the dialysis system for 24 hours.

Taken together these data indicate that the dialysis system used here appears to be a good enough method to allow small volume culture of *Brugia malayi*. In addition it should be possible to use both lifespan and Mf release in culture as robust phenotypes for preliminary use in RNAi experiments using genes that would be expected to significantly affect them compared to controls.

Fig 5.6 Effect of small volume culture methods on lifespan of adult female *B. malayi* in culture

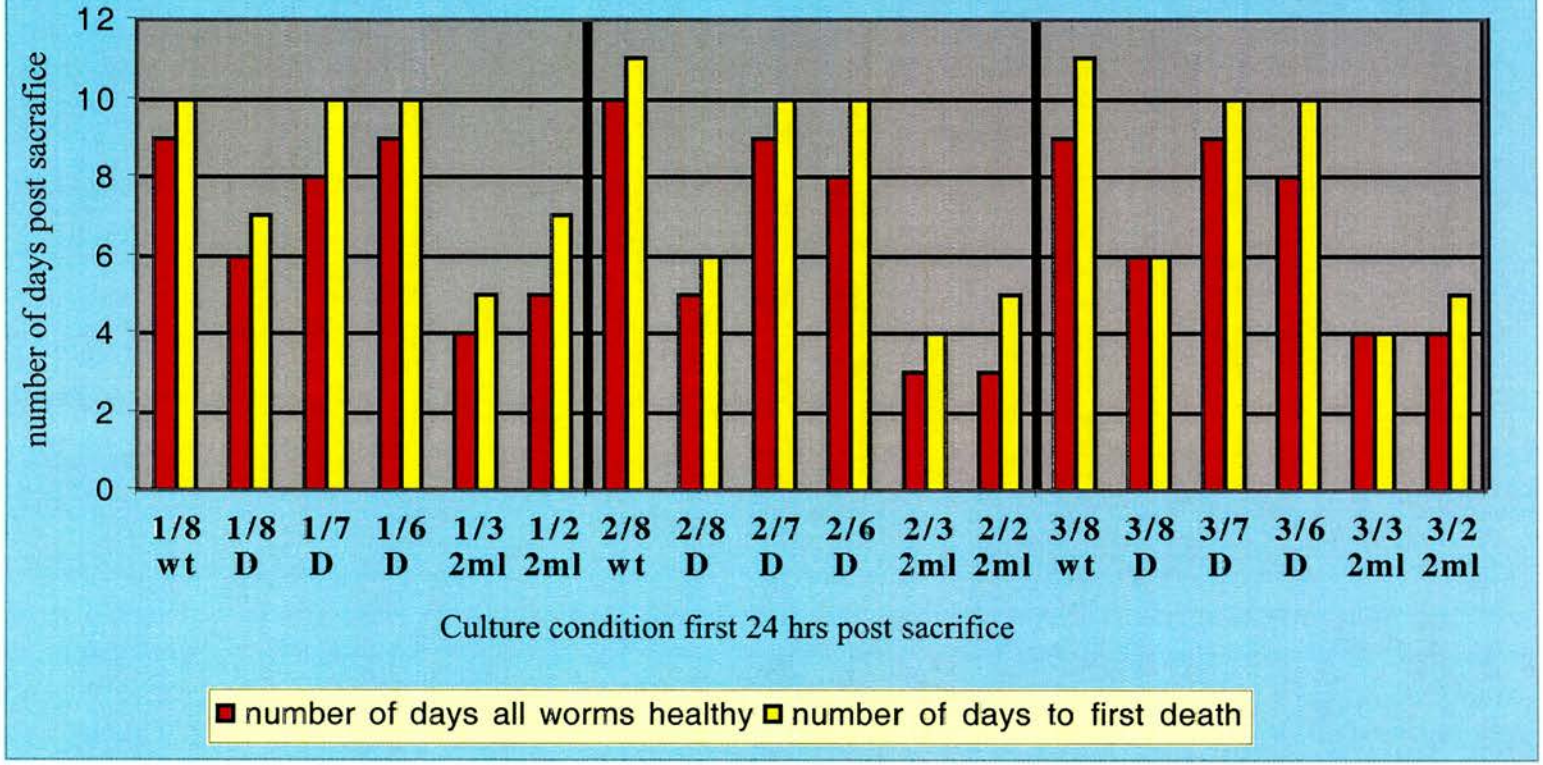


Fig 5.6 Effect of small of small volume culture methods on lifespan of adult female *B. malayi* in culture

Three groups of 34 adult female parasites taken from three different jird hosts (1,2 and 3) were split into groups and cultured and in three different culture conditions. Parasites were grouped on the basis of their microfilarial release such that each group had approximately the same average Mf release/worm/hour. Worms were either cultured in standard conditions (wt), in the dialysis system (D) or in 2ml of media (2ml) as described above in groups of 8, 7, 6, 3 or 2 as indicated on the axis. They were maintained in these culture conditions for the first 24 hours post sacrifice and then moved to standard culture conditions (or maintained in them). Worm health was then observed by assessing motility, which is normally characteristically very vigorous in culture until worms become unhealthy. The day post sacrifice on which the first worm in a group was scored as unhealthy was recorded as well as the day on which the first worm in a group was scored as dead (no motility and floating).

Culture in small volume culture for 24 hours post-sacrifice significantly decreases the lifespan of parasites as does culturing 8 worms in the dialysis system. However culturing less than eight worms in the dialysis system does significantly decrease lifespan. This suggests that culturing *B. malayi* in the dialysis system will not significantly decrease the lifespan and health in culture and is therefore a suitable method for small volume culture for use in the development of RNAi technology in this parasite.

Fig 5.7 Comparison of microfilarial release after standard vs dialysis culture for 24 hours post-sacrifice

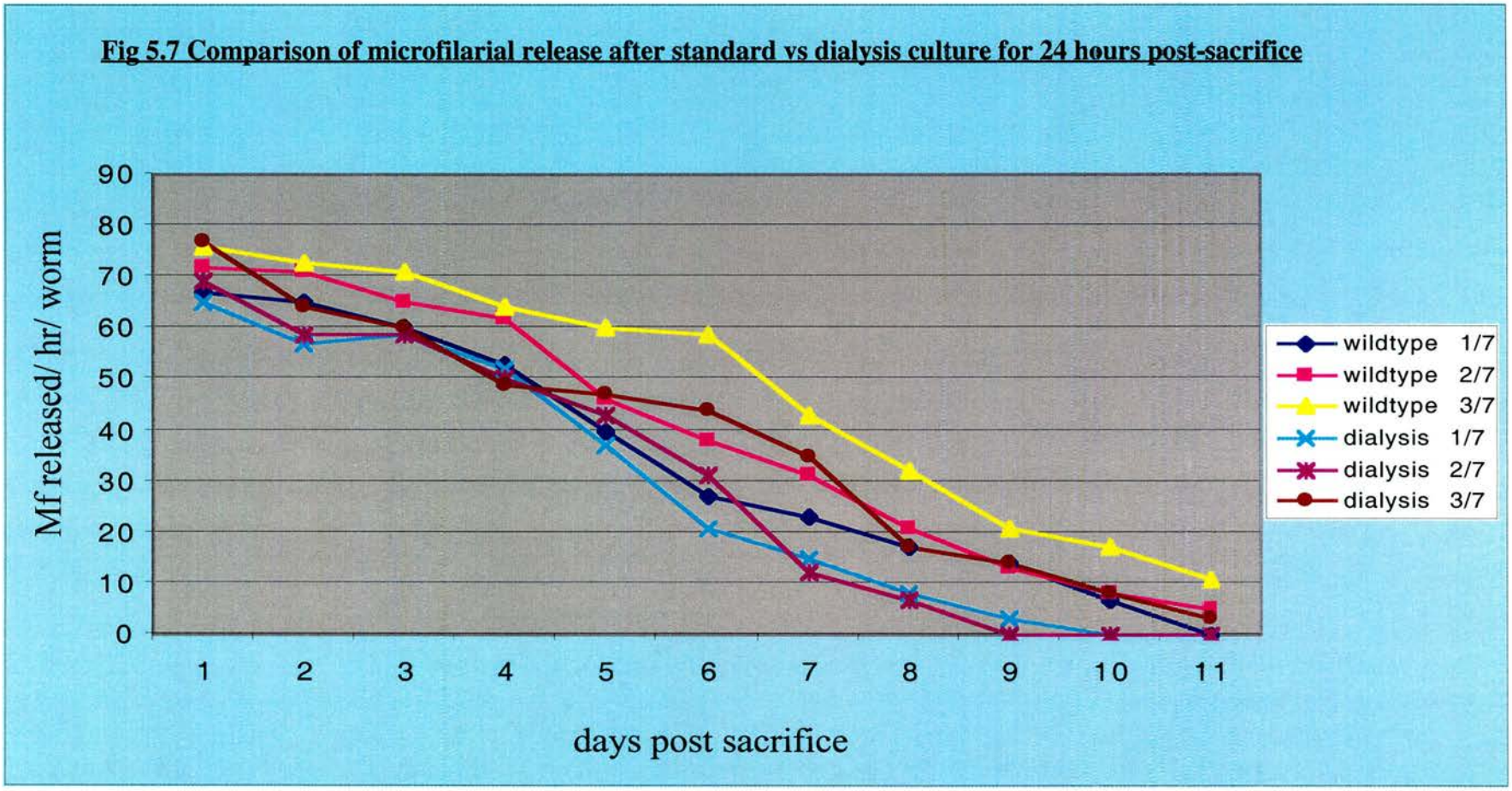


Fig 5.7 Comparison of microfilarial release after standard vs dialysis culture for 24 hours post-sacrifice

Microfilarial release was assessed for groups of parasites cultured in either standard conditions (wt) or the dialysis system (D) for the first 24 hours post-sacrifice from 3 different jird hosts (1, 2 and 3). These jirds correspond to the same jirds for which health and lifespan were assessed (Fig 5.6). Initial groupings were made such that the group in the dialysis system and the group in standard culture conditions had approximately the same average Mf release per hour per worm.

Mf release from worms maintained in the dialysis system was consistently lower than that of worms maintained in the dialysis system. However, despite this drop in Mf release the difference is not large enough to preclude Mf release as a possible measure of the effect of RNAi and suggests the effect of dialysis culture on health as measured by Mf release is only slight. Experiments using Mf release as a measure of the effect of RNAi can be compared to that of worms maintained in the dialysis system

5.9. Choice of target genes for RNAi in *Brugia malayi*

Three genes have been chosen as preliminary targets for RNAi in *Brugia malayi*. Two of these, RNA polymerase II gene (*Bm-ama-1*) (Aboobaker and Blaxter unpublished data) and β-tubulin (*Bm-tub-1*) (Helm et al., 1989), have been chosen because they are expected to be essential for adult parasites to survive. Reduction in transcript levels of these genes in culture would be predicted to significantly reduce worm health in culture and perhaps even result in death of adult worms. The third gene is sheath-protein-1 (*shp-1*) known to be a component of the Mf sheath (Selkirk et al., 1991b). The transcript is highly abundant and expressed only in adult females and as such represents an ideal candidate for RNAi in the system described here (Selkirk et al., 1991b). Loss of mRNA transcript, leading to loss of protein might affect the structure of the Mf sheath. This phenotype could be observed in the microfilaria released in culture by adult female parasites, proving the usefulness of RNAi for studying genes involved in the development of Mf and providing more insight into the function of this particular gene

In preparation for using this gene coding fragments of each gene were cloned and used as templates for PCR reactions with T7 and T3 primers to produce template DNA for in-vitro transcription reactions. RT-PCR reactions on panels of individual worms were performed for each gene to show that the optimised PCR conditions gave consistent results (Table 5.1). On rare occasions some worms had apparently lower or higher expression levels for all genes (two genes were assessed for each single worm PCR), this could be due to either worm size or occasionally better or worse reverse transcription reactions. However, as this was consistent for all genes it does not affect the use of single worm RT-PCR as a convenient method for assessing transcript levels.

5.10. RNAi of *Brugia malayi* β-tubulin and RNA polymerase II large subunit genes results in reduction of transcript levels and death of adult female worms in culture

Before placement of adult female parasites in the dialysis system Mf release counts were approximately normally distributed. From this population 35 adult female worms from the

middle of the distribution were pooled into four groups of seven worms each. These groups were chosen so that each group had the approximately same average Mf release count. The first group of worms was placed in standard culture conditions for the duration of the experiment, the second group was placed within the dialysis system for the first 24 hours and the remaining three groups were soaked in dsRNA corresponding to one of the genes described above.

RNAi by soaking was performed on groups of seven adult female parasites using the dialysis system described here in concentrations of dsRNA in excess of 3mg/ml. Worms were removed after 10, 14, 17, 20 and 24 hours in the dialysis system to assess transcript levels by single worm RT-PCR.

Control worms showed normal transcript levels at all time-points and were observed to be healthy throughout the 24-hour period with only a small reduction in motility (Fig 5.8). This reduction in motility was reversed when parasites were placed back into standard culture conditions. Transcript levels of both *Bm-ama-1* and β-tubulin were reduced during the 24-hour period by soaking in complementary dsRNA.

5.10.1. RNAi of Bm-ama-1

The transcript level of *Bm-ama-1* started to diminish between 14 and 17 hours in the dialysis system (Fig 5.8). At the 17-hour time-point the remaining worms appeared unhealthy and showed reduced motility compared to control worms. By comparison with the data obtained from PCR optimisation experiments we can estimate that this drop represents loss of two thirds of normal transcript levels. The transcript level had further reduced at 20 hours and worms were observed to have further diminished motility. At this time a significant drop in *Bm-tub-1* transcript (assessed as an internal control) was also observed although levels of *shp-1* dropped only very slightly. It is possible that the loss of RNA polymerase II protein causes a reduction in the amount of new transcripts being produced and this is reflected in the reduction of the other transcripts observed here. The reduced levels of transcript could also be a secondary effect of general breakdown in cellular mechanisms caused by reduced transcript levels or other essential genes require for parasite survival. After 24-hours the remaining two worms were scored as dead as

they showed no motility and failed to recover on transfer to normal culture conditions. RT-PCR of the worm collected at this time point did not detect any transcript for any of the genes.

5.10.2. RNAi of Bm-tub-1

The transcript level of *Bm-tub-1* started to diminish between the 10 and 14-hour time points and continued to drop to just detectable levels after 20 hours, with the majority of the transcript lost after 17 hours (Fig 5.8). However, the transcript levels of the non-target genes also dropped to just detectable levels after 20 hours. *Bm-ama-1* levels dropped after *Bm-tun-1* levels and before *shp-1* levels. After 20 hours the remaining worms were observed to have very low motility and were close to death, after 23 hours only one worm remained alive and this worm was dead by the 24-hour time point. RT-PCR on the worm collected for the 24 hour time point failed to detect transcript for any of the genes. This not surprising as mRNA would be likely to degrade rapidly on worm death. Similar to what was observed for RNAi of *Bm-ama-1* it appears that loss of β-tubulin transcript most probably leads to critical reductions in protein levels that quite rapidly leads to worm death in culture. The drop in transcript levels of the non-target genes, and possibly the further reductions in *Bm-tub-1* levels after 17 hours, could have resulted from a general breakdown of cellular mechanisms due to reduced levels of *Bm-tub-1* protein.

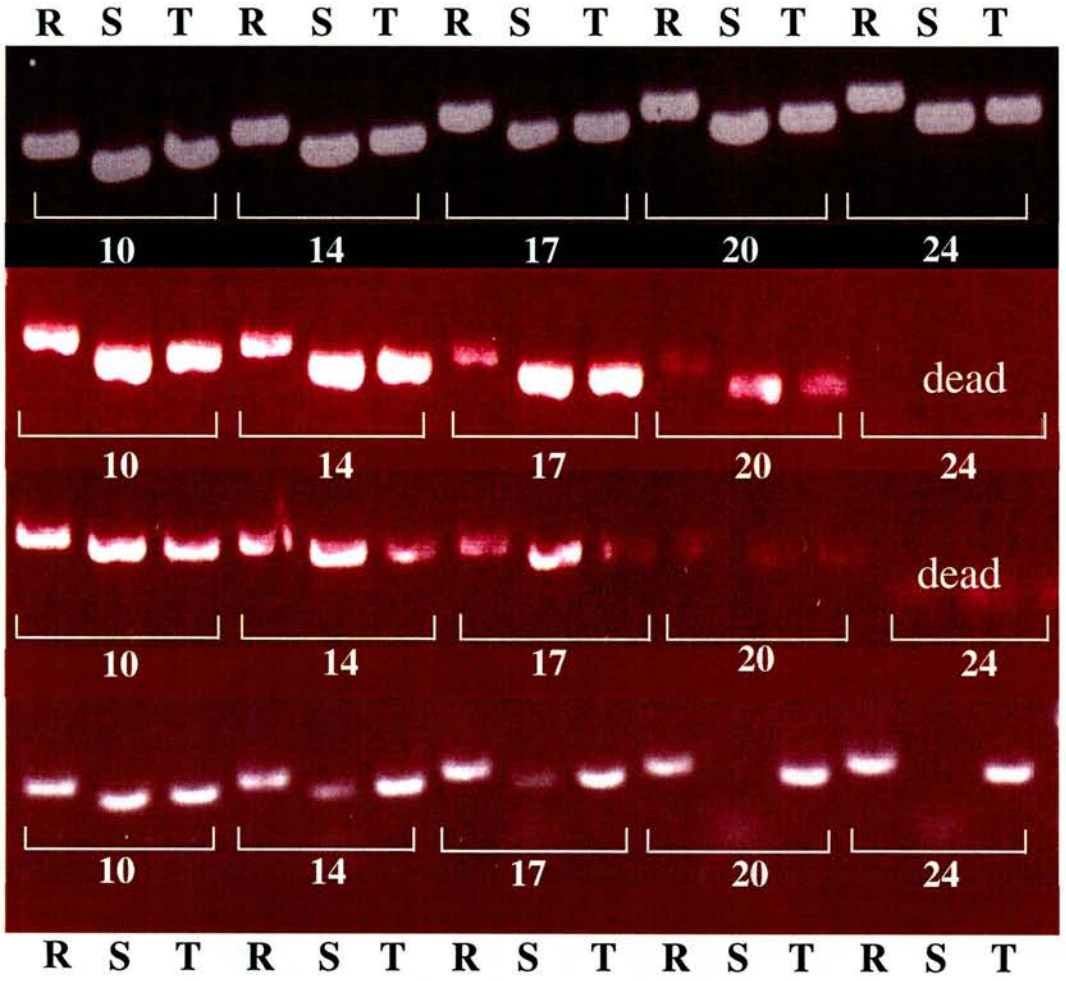
5.10.3. RNAi for proof-of-principle putative drug target candidates

The RNAi of two essential housekeeping genes has similar results, rapidly resulting in worm death in the culture system used here. Both of the proteins that are coded for by these transcripts perform essential functions in every cell of the parasite and it is likely that not every cell need be adversely affected to induce worm death. Within individual cells it is not clear how the observed transcript loss translates into a drop in protein levels. It may be that only a relatively small drop in protein levels of either RNA polymerase II or β-tubulin is toxic to individual cells, resulting in cell death or improper functioning.

Occurring on a tissue level this would be likely to result in death in the culture system, which is likely to be a less than ideal substitute for the host lymphatic or peritoneal environments.

These observations suggest that RNAi could test putative drug target candidate genes required for parasite survival under these conditions.

Fig 5.8 Single worm RT-PCR to assess transcript levels in control and treated worms



Control worms

Time after Exposure (hr)

RNA pol II dsRNA

Time after Exposure (hr) **3.6mg/ml**

ß-tubulin dsRNA

Time after Exposure (hr) **3.7mg/ml**

sheath protein-1 dsRNA

Time after Exposure (hr) **3.5mg/ml**

Fig 5.8 Single worm RT-PCR to assess transcript levels in control and treated worms

Groups of seven adult female *Brugia malayi* were exposed to high concentrations of dsRNA for 24 hr post-sacrifice. Worms were grouped, after their Mf release frequencies had been measured, so that each group had approximately the same average Mf release frequencies. After 10, 14, 17, 20 and 24 hours individual worms were removed from each group, including the control group and stored for later RNA and first strand cDNA preparation.

First strand cDNA was prepared for each worm and the level of *Bm-tub-1* (T), *Bm-ama-1* (R) and *shp-1* (S) transcript assessed for each worm as described above. Transcript levels of all genes were normal for control worms not exposed to dsRNA. Target specific reduction in transcript level was observed for worms taken from the treated groups (described in the text). After 24 hr the remaining worms treated with *Bm-tub-1* and *Bm-ama-1* dsRNA were dead.

5.11. RNAi of *Brugia malayi* Mf-22/shp-1 gene results in specific transcript reduction and the production of "short" microfilaria in culture.

RNAi of the *shp-1* gene did not result in worm death although the worms remaining at the end of the 24 hours in the dialysis system showed markedly reduced motility on their return to standard culture conditions compared to control worms. Reduction in *shp-1* transcript levels began between 10 and 14 hours and continued to decrease until they were no longer detectable at 20 and 24 hours using the optimised conditions for single worm RT-PCR (Table 5.1). Transcript was detectable however when PCRs were performed with 35 cycles at both of these time points. It is likely that the optimised single worm PCR is detecting a transcript that is very highly expressed and does not detect the reduced levels induced by RNAi. No transcript reduction was observed for either of the non-target genes (in contrast to experiments when they were targeted). This suggests that, unlike *Bm-ama-1* and *Bm-tub-1*, *shp-1* is not essential for parasite survival in culture in the short term. As the protein from this maternal/embryonic transcript appears only in the Mf sheath and appears to have no function in adult parasites this is not surprising. In order to further investigate the function of the *shp-1* and the dynamics of RNAi in this system the remaining two worms at the end of the 24-hour period were moved to standard culture conditions.

Mf release counts were performed on both control and *shp-1* RNAi worms immediately after removal from the dialysis system, and again after a further 24 and 48 hours. The Mf release counts for the *shp-1* RNAi worms were lower than those of control worms, which in turn were only slightly lower than the counts for worms that were not placed in the dialysis system (Table 5.3). This difference is likely to be due to loss of the *shp-1* transcript (discussed below).

Released Mf were fixed and Giemsa stained to observe the phenotypes of their sheaths. Half of the Mf released by *shp-1* RNAi worms in standard culture after 48 hours had malformed sheaths (Table 5.3, Fig 5.9). This was characterised by apparently complete embryogenesis through to normal looking Mf coiled up within sheaths that had failed to elongate to greater or lesser extents. This spectrum of phenotypes suggests that loss of *shp-1* transcript in the gonad of adult female parasites interrupts normal elongation of the embryo eggshell into the Mf sheath. It is likely that normal Mf released in culture

represented those far enough along in development not to be affected by loss of *shp-1* transcript, with increasing degrees of severity for the Mf released later. It is not clear with the reduction in total release is due primarily to loss of *shp-1* transcript or whether this due to some secondary factor that interruption to Mf sheath elongation causes. Some aberrant Mf release was also observed from control worms and included in this category. However these were mostly examples where embryogenesis was not complete and appear to occur due to failures in normal processing in this culture system.

SHP-1 belongs to a group of proteins either known to or supposed by analogy to incorporate into the Mf sheath structure (Conraths et al., 1993; Hintz et al., 1994; Hirzmann et al., 1995; Selkirk et al., 1991b). *shp-1* itself is known to be an entirely maternally/embryo derived transcript as the transcript is absent from mature Mf (Selkirk et al., 1991a). Work on the *shp-1* from *Litomosoides carinii* (a closely related filarial nematode) indicates that transcript is present in large amounts within embryos during development while no protein is detected (Conraths et al., 1993). During development and elongation of the Mf sheath, translation of stockpiled transcript occurs and mature SHP-1 protein is moved to the eggshell/ Mf sheath. In both species its structure suggests that it may contribute to the flexibility of the Mf sheath (Conraths et al., 1993; Selkirk et al., 1991b).

To look at this possibility collected Mf were observed for motility in culture. Mf collected from *shp-1* RNAi adults that appeared normal and those collected from control worms had apparently equal motility and flexibility by comparison by eye under light microscopy. Abnormal Mf that were not elongated had either no or very low motility, accept for rare individuals that had undergone almost complete elongation that showed intermediate erratic motility as they attempted to fully elongate within their sheaths. The lack of motility in all but the most mildly affected could just be a function of Mf health and not related to lack of flexibility of the eggshell/sheath structure. To test this Mf collected from the *shp-1* RNAi worms were ex-sheathed to see if they could regain motility. However, this experiment was only an afterthought and Mf collected from the last time point were not healthy enough to survive the ex-sheathment protocol once they had already been centrifuged once for counting purposes. This was also true of the Mf collected from control worms so is unlikely to reflect lower health of *shp-1* RNAi Mf. Nonetheless it should be possible to look at this again when repeating *shp-1* RNAi.

To see if the effect of *shp-1* RNAi was maintained over the 48-hour period for which the worms were observed single worm RT-PCR was performed on the remaining worms. For both worms *shp-1* transcript was not detectable using the standardised optimised single worm RT-PCR, were as it was detectable in control worms. Again, transcript was detected when PCR reactions were performed with 35 cycles as described above for worms taken from the 20 and 24 hour time-points. This suggests that the transcript level remains reduced even after worms have been removed form the dsRNA.

Table 5.3 Mf release of *shp-1* RNAi and control worms 48 hours post treatment

48 hours post treatment	Number of normal Mf released/worm/hr	Number of short Mf released/worm/hr	Total number mf released/worm/hr
shp-1 dsRNA treated worms	10	9	17/19
Control worms	29	2	31/31

Table 5.3 Mf release of *shp-1* RNAi and control worms 48 hours post treatment

Mf release was measured for control and *shp-1* dsRNA treated worms 48 hours after being removed from the dialysis culture system by light microscopy (first figure in last column) and the phenotypes and release assessed by counting Giemsa stained Mf on slides. The Mf release from treated worms was 17/worm/hour, less than 30/worm/hour observed for control worms. In addition of the Mf released from treated worms 8/worm/hr had failed to fully elongate within their sheaths (Fig 5.9). These were scored as being 'short' and varied in severity from those that had partially elongated and those that had failed to elongate. Some of the Mf released by control worms also scored as not normal. However, although they have been included in the 'short' category here, they were characterised by incomplete embryogenesis rather than incomplete sheath elongation.

Mf release/worm/hour was calculated as described above in the section "Quantification of microfilarial release in culture" and is represented by counting the Mf released by the remaining worms over three hours (2 control worms and 2 shp-1 dsRNA treated worms). The raw data for these counts is presented in Appendix 4

Fig 5.9 Treatment with shp-1 dsRNA leads to the production of 'short' Mf

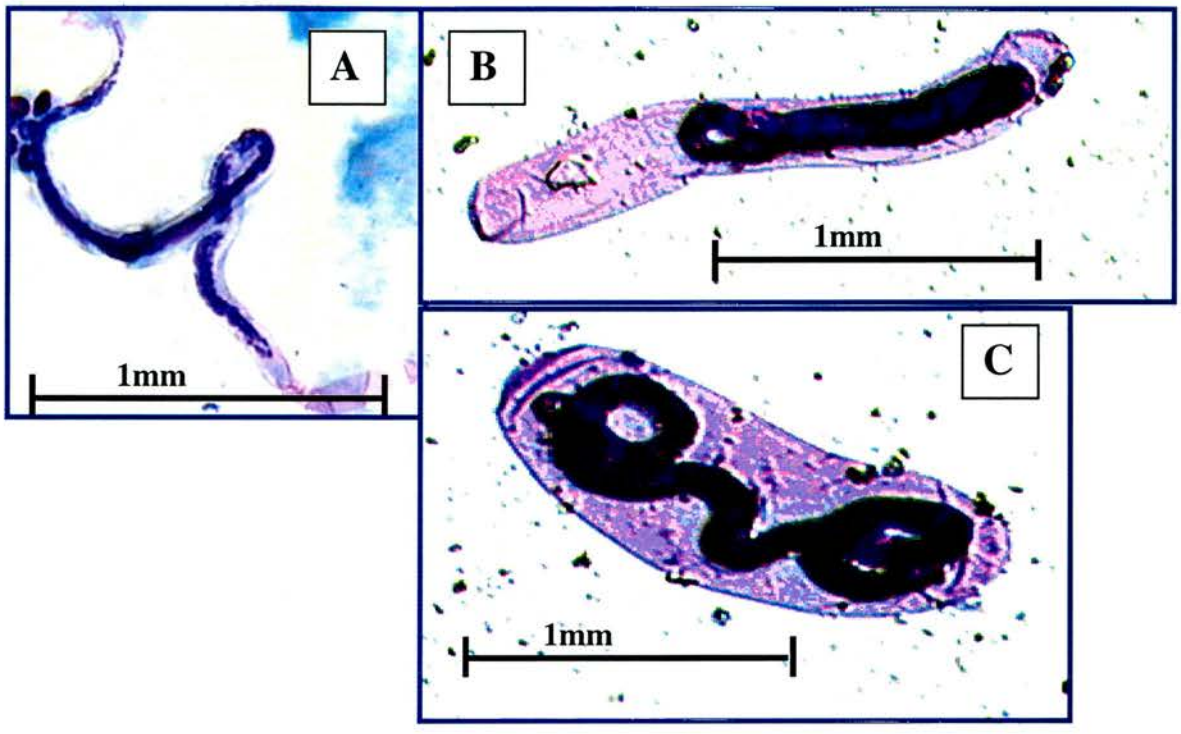


Fig 5.9 Treatment with shp-1 dsRNA leads to the production of 'short' Mf

Approximately half the Mf released by shp-1 dsRNA treated adult female parasites had a charcateristic 'short' phenotype (B and C), rather the normal elongated sheath (A). The 'short' phenotype had various grades of severity from Mf from mild to intermediate (B) and severe (C). Embryogenesis seems to be complete as Mf appear to be fully developed, but they are unable to elongate as the sheath has failed elongate around them. This probably due to loss of the Shp-1 protein which is normally incorporated into the Mf sheath and is hypothesised to allow flexibility.

5.12. RNAi in *Brugia malayi*

5.12.1. The need for further characterisation and development of RNAi in B. malayi

From the preliminary data described here we can conclude that the mechanism of RNAi works in the filarial parasite *B. malayi*. However, the dialysis system for RNAi in culture has some limitations to the spectrum of genes that can be usefully investigated with the phenotypic readouts available. Perhaps the biggest drawback of the system is that only adult parasites are amenable to the technique and thus only genes with an adult function will produce useful data when targeted. One additional set of genes amenable to study in this system are those required for proper formation and release of Mf in culture, potentially a very important set of genes for finding novel drug targets. Another drawback is that the culture system described here is inherently artificial. The host-parasite interaction is of key importance in studying parasite biology, but many of the parasite genes involved in this process are unlikely to be essential for parasite survival in culture. The sets of genes that fall into this category include parasite derived immune modulators required for establishment of long term parasite survival in the lymphatics, enzymes required to survive direct insults from the host environment and possibly some of those genes required to make use of host derived nutrients (Maizels et al., 2001). This last set of genes might also be required for survival in culture conditions and could be amenable to study in this system.

From this discussion it is clear that further characterisation of this system and the dynamics of RNAi in *B. malayi* is required to make it into a more powerful tool for studying parasite gene function (discussed below). This development includes a clearer understanding of what transcripts can be targeted and also the possibility of employing other assays, including finding ways to link RNAi to well established biochemical and immunological systems for looking at gene function. If this proves to be possible then both the long term and short-term host parasite interactions involved in parasite establishment and survival might be illuminated by experiments that utilise RNAi.

5.12.2. Investigating the link between transcript loss and protein levels: which tissues are targeted and for how long by RNAi in B. malayi

The experiments described here have not yet addresses the effect of RNAi on the protein levels of targeted genes. All the parasite samples for which transcript level has been assessed can also be assessed for protein levels of targeted genes by Western blot analysis. Antibodies for all three genes are available and these will allow protein levels to be compared to the transcript levels described here (Helm et al., 1989; Selkirk et al., 1991a; Seydoux and Dunn, 1997).

We would expect to see drops in levels of all proteins by Western blot analysis. The non-specific transcript loss detected for RNA polymerase II and β-tubulin RNAi probably results from general degeneration due to a critical drop in levels of these essential proteins. However, for this effect to occur it would not be necessary for the whole worm to be affected by RNAi. It is possible that significant drops or complete loss of transcript in just one or a few critical tissues/cell groups could eventually lead to worm death in culture. The transcript loss of RNA polymerase II is a specific effect up to 17 hours and then the other transcript levels begin to drop. Western analysis should reflect this by revealing some drop in protein levels by this time. This analysis will not however be informative about which tissues are affected by RNAi. For this an immuno-localistaion studies will be needed in future experiments with genes expressed in all cells, on serial cross-sections of whole parasites to establish which tissues are targeted preferentially or if all tissues are targeted equally. This question can also be addressed by looking at other genes with known tissue specific expression patterns to see if transcript and protein levels drop.

From the *shp-1* transcript data we can assume that RNAi efficiently targets the gonad. This gene is exclusively expressed within developing embryos and Mf with the protein being transported as 22 kDa monomers to the sheath structure of Mf where it is detected as being cross-linked in multiples of 22 kDa (Conraths et al., 1993; Selkirk et al., 1991a). Western analysis should reveal a relative reduction in the pool of 22 kDa monomers available for transport to the Mf sheath, resulting the "short" phenotype observed here. More significantly, work on *shp-1* in the closely

related filarial nematode *Litomosoides carnii* shows that the transcript is within developing embryos (Conraths et al., 1993). This is likely to be the case for *Brugia* and indicates that RNAi is able to affect developing embryos. This could be a very useful tool for studying gene function, including the function of genes involved in developmental processes.

Another important question that must be addressed is how long the RNAi effect is retained for. Many immunological models of filariasis require that adult worms be re-implanted into hosts so that he effects on the host immune response can be investigated. Ideally RNAi cold be used to target the expression of possible immune regulatory candidates produced by the parasite (Falcone et al., 2001). Some of these are already well characterised, such as the filarial macrophage migration inhibitory factors thought to play an important role in inducing the activation of special subsets of host immune cells. However, to look at the effects of infections without these genes any RNAi effect would have to be maintained for several weeks. It has not been assessed yet how long the RNAi effect is maintained for in *B. malayi.* The data from *shp-1* RNAi worms indicates that in culture the effect remains for at least 48 hours. This is encouraging but, but future experiments will have to involve the RNAi of worms, introduction back into hosts by implantation and then assessment to see for how long RNAi is effective. It not clear how long worms will survive for in the host after being through the dialysis system, but normal worms can survive for several weeks in the peritoneal cavity of mice.

5.13. Prospects for RNAi in *B. malayi* and other parasitic helminths

It appears that RNAi may be a major breakthrough for studying gene function in parasites. The approach used here should be transferable to other parasites, especially ones that are smaller than *B. malayi* and are likely to pose less technical difficulties. If RNAi can be applied correctly it may be possible to use it as a tool for discovering parasite specific genes essential to their survival. If this proves possible it not only provide a valuable insight into parasite biology but identify proteins that may be ideal drug target candidates for the development of future therapies. The study of

parasitic helminths is being linked more and more with the study of the immune system. RNAi could prove to be the breakthrough tool for analysing the effects of parasite products on the modulation of the host immune response *in vivo*.

5.14. Using RNAi to study Hox gene function in *Brugia malayi*

As previously discussed at the beginning of this chapter RNAi has been applied to a number of animals previously not amenable to functional genetic analysis. The technique of RNAi in *Brugia malayi* will also allow the possibility of studying the function of conserved developmental genes, such as the Hox genes.

The targeting of the *shp-1* transcript in particular suggests that embryonic transcripts may be particularly amenable to RNAi. Temporal expression pattern analysis of the *Brugia malayi* Hox genes shows that they are all expressed in developing embryos and so they should be amenable to RNAi. Elucidating the function of the Hox genes in a nematode only distantly related to *C. elegans* would provide a better understanding of Hox gene evolution in the Nematoda.

While detailed analysis of the *Brugia malayi* Hox gene RNAi analysis may not be possible to the same extent achieved for *C. elegans* some major questions may be simply addressed. Firstly, as previously discussed, in *C. elegans* it is known that three Hox genes have only post embryonic function in contrast to both vertebrate and arthropod Hox gene complements which are all involved in embryonic patterning. It should be possible to test if loss of individual *Brugia malayi* Hox gene transcripts has any affect on embryogenesis in this nematode. Secondly it should be possible to test the function of the two extra Hox genes that *Brugia malayi* possesses in comparison to *C. elegans*. One prediction might be tested is that the evolution of solely post-embryonic functions might be an intermediate stage in Hox gene loss that is observed in *C. elegans*.

5.15. Acknowledgements

I would like to say an extra special thank you to my supervisor Dr Mark L. Blaxter for allowing me to conduct these experiments and continuing to have faith in me even when much grant money was spent and no data appeared. The work here not only pays tribute to his open mindedness in the lab but also his willingness to encourage and believe in the talents of those around him beyond the point when more "sensible" scientists wouldn't. At the time of writing this work remains without any direct funding despite a number of grant applications being submitted to suitable schemes.

Without the enthusiastic support of Professor Rick Maizels and his generosity with the *Brugia malayi* lifecycle no progress would have been possible. For this and for his encouragement throughout I am very grateful.

Without the excellent technical assistance I received from Mrs Yvonne Harcus, early on the piece, in parasite lifecycle maintenance and parasite culture I would have remained mostly clueless. Her patience, understanding and hugs have helped me through the most frustrating periods of this work. When strings of 24 hour vigils, occasionally back to back, resulted in only minimal progress she was the first to ask me how I was in the morning and remind me not to take it all too seriously.

CHAPTER 6: MATERIALS AND METHODS

In this section the materials and methods used for experiments described in the preceding chapters are described. These include all the general molecular methods, a detailed description of the primers and PCR conditions used in the degenerate PCR screen, and a description of the molecular phylogenetic methods used in chapter 4. The materials and methods particular to the development of RNAi in *B. malayi* are described in chapter 5.

6.1. Preparation of gDNA from nematodes

Genomic DNA was prepared from nematodes collected and frozen at -80°C or prepared from materials left over from RNA extraction (see below). DNA from primary samples was prepared by a standard phenol:chloroform extraction procedure as follows. Frozen Nematodes were homogenised in sterile 1.5 ml tubes in 500 µl Worm Lyis Buffer 2 (110 mM NaCl, 110 mM Tris-cl. PH 8.5, 55 mMEDTA, 1.1% SDS, 1.1% 2, mercaptoethanol) and 0.1 mg/ml proteinase K (Sigma, UK). After homogenisation the tube was incubated at 65°C for 30 minutes with occasional very slow inversion.

After proteinase K digestion 500 µl of phenol was added and mixed by very slow inversion for 5 minutes followed by spinning in a microcentrifuge for 10 minutes at 12, 000 rpm. The lower organic phase was then carefully removed and 500 µl of phenol:chloroform was added and mixed by very slow inversion and spun as described above. After spinning the supernatant (top phase) was removed with a wide bore pipette tip and placed in a fresh 1.5 ml tube. This volume was then extracted with 500 µl of chloroform and mixed and spun as described above, The supernatant was again moved to a fresh tube with a wide bore pipette tip.

DNA was precipitated by adding 1 ml of 100% ethanol at –20 °C and spinning at maximum speed for 15 minutes to pellet the DNA. The resulting pellet was washed twice with 500 µl of 70% ethanol, mixing gently and spinning after each wash.

Finally the pellet was washed with 100 µl of 95% ethanol, allowed to air dry and then resuspended in 50 µl of sterile water or 1x Tris-EDTA.

Preparation from samples used first for RNA extraction was as follows. After complete removal of the aqueous phase containing RNA (see below) 300 µl of 100% ethanol was added to the remaining extract and left at room temperature for 3 minutes followed by spinning at 2.000g for 5 minutes at 4°C. The supernatant was then removed and the DNA pellet washed twice in 1 ml of 0.1 M sodium citrate in 10% ethanol leaving the pellet in the wash solution for 30 minutes at roo temperature and centrifuging at 2.000g at 4°C in between washes. The pellet was then air dried and resuspended in 600 µl of 8 mM NaOH and spun at maximum speed for 10 minutes to pellet residual material. The supernatant containing the DNA was then removed to a fresh tube and precipitated as described above and resuspended in sterile water or 1x Tris-EDTA.

All resuspended gDNA samples were quantified using a DNA spectrophotometer (Genequant, ABgene, UK)

6.2. Preparation of RNA and first strand cDNA from nematodes

6.2.1. RNA extraction from nematode samples

Nematode samples were frozen at –80°C in Trizol (Invitrogen Life technologies) and then thawed on slowly on ice. Samples were homogenised in sterile 1.5 ml tubes or ground using a percussive disruptor and a hammer with regions of contact with the sample pre-cooled in liquid Nitrogen. Samples were then incubated at room temperature for 10 minutes in 1 ml of Trizol before adding 200 µl of chloroform and shaking vigorously for 15 s. Samples were then left at room temperature for a further 5 minutes before spinning at maximum speed in a microcentrifuge for 15 minutes at 4°C. The upper colourless aqueos phase was then removed and placed in a fresh tube. Total RNA was precipitated using 500 µl of isopropanol and incubated on ice for 30 minutes. The sample was then spun at maximum speed in a microcentrifuge for 15 minutes at 4°C. The RNA pellet was then washed once with 75% ethanol,

mixed by vortexing and spun at 7,500 g for 5 minutes at 4°C. RNA pellets were then stored under 100% ethanol at -80°C or resuspended in DEPC treated water for quantifying with a DNA spectrophotometer (Genequant, Abgene) and subsequent use in reverse transcriptase reactions. The organic phase was maintained to allow the extraction of DNA (see above).

6.2.2. Production of first strand cDNA

First strand cDNA was synthesised using the ProSTAR Ultra High Fidelity RT-PCR System and the protocol therein. Briefly 500 ng of total RNA was used in a 50 µl reaction containing the following provided reactants: 3 µl of oligo dT primer (100 ng/µl), 5 µl dNTP mix (40mM), 5 µl 10x StrataScript RT buffer, 2.5 µl StrataScript RT enzyme (20 units/µl), 2 µl Rnase inhibitor (2 units/µl) made up to a volume of 50 µl with Rnase free water. The reaction mix was incubated at 65°C for 5 minutes and allowed to cool at room temperature to allow oligo dT to anneal and remove secondary structure before addition of RT enzyme and Rnase inhibitor. The complete reaction was then incubated at 42°C for 1 hr. First strand cDNA was stored at -20°C before use in PCR reactions.

6.3. Primers and PCR conditions for degenerate PCR and subtractive screening

6.3.1 Primers and primer design

The primers used to amplify nematode Hox genes can be broadly split into two categories. Those that are "universal" and would be expected to amplify multiple Hox and/or Hox related genes and those that are more specific and expected to amplify specific genes or groups of genes. Table 6.1 contains the sequences of all the primers used in this study. Some primers were the same as those used in previous studies (de Rosa et al., 1999 and refeences therein). Primers used here for the first time were designed by aligning orthologous groups of proteins, identifying conserved motifs and designing primers to them with the aid of the BLOCKS

algorithm family, which will design degenerate primers (http://blocks.fhcrc.org/blocks/process_blocks.html). These primers were than modified by eye to reduce excess degeneracy and introduce split codons at their 3' ends.

Table 6.1 Degenerate PCR primers

Primer class	Primer name	Primer sequences
Universal forward primers	SO1	GARYTNGARAARGARTT
	B01F1	GCTTCTAGARYTNGARYTNGARAARGARTT
	ELEKEFK	GCTCTAGARYTNGARAARGARTTYA
	LELEK	CGGATCCCTIGARCTIGARAARGA
	LELEKE	TACCAGACCYTNGARYTNGARAARGA
Universal reverse primers	KIWFQN	CATAGCCGCRTTYTGRAACCADATYTT
	BO2R1	GGAATTCRTTYTGRAACCADATYTT
	BO2R2	GGAATTCRTTYTGRAACCANACYTT
	WFQNRRM	CATWCKWCKRTTYTGRAACCA
	S02	CKNCKRTTYTGRAACCA
	RQIKIWF	RAACCARATYTTRATYTGNG
ceh-13 group forward primer	TNFTTHQL	ACWAACACWTTYACWACWCAWCARYT
lin-39 group forward primers	AYTRNQY	GCWTACACWMGNGAYCARTA
	RGEKRQR	MGAGGWGAGAARMGNGARMG
lin-39 group reverse primers	FQNRRMKH	RTGYTTCATWCKWCKRTTYTG
Central orthology group forward primers	KRGRQTY	AARMGAGGWMGAGARACWTA
	TYTRYQT	ACWTAYACWMGWTAYGARAC
	GRQTYTR	GGYMGAGARACWTAYACWMG
mab-5 group forward primers	TYSRTQT	ACWTAYTCYGGYACWGARAC
	TRQTYSRT	ACWMGAGARACWTAYTCYMGAAC
mab-5 group reverse primer	QNRRMKH	YTTCATWCKWCKRTTYTG
egl-5 group forward primers	QTYQRYQT	GARACWTAYGARMGATAYGARAC
	RQTYQRY	MGAGARACWTAYGARMGATA
	KGRQTYQ	AARGGYMGAGARACWTAYCA
	GRQTYQR	GGYMGAGARACWTAYGARMG
Posterior group genes forward primers	PYTKYQT	CCWTAYACWAARTAYGARAC
	RKKRKPY	MGAAARAARMGAAARCCNTA
	RKKR(K/R)PY	MGIAARAARMGIMRNCCNTA

Table 6.1 Degenerate PCR primers

Degenerate PCR primers from other studies are in red all other primers were designed in this study. Primer design is described in the main body of the text (see above). All primers are in a 5' to 3' orientation.

6.3.2. Degenerate PCR conditions

A number of different PCR conditions were tried and optimised by performing PCR against gDNA and cDNA prepared from *C. elegans* and *B. malayi*. Two different PCR cycles based on touchdown PCR produced consistently good results, and one was used for amplifying from cDNA and the other for amplifying from gDNA. The programme used for cDNA was 95°C for 1min for 1 cycle, (95°C for 15 s, 50°C to 40°C decreasing by 1°C each cycle for 20 s, 72°C for 45 s) for 10 cycles, (95°C for 15 s, 44°C for 20 s, 72°C for 45 s) for 25 cycles, 72°C for 5 minutes, 4°C 5 minutes. The programme used for gDNA was 95°C for 1min for 1 cycle, (95°C for 15 s, 50°C to 40°C decreasing by 1°C each cycle for 20 s, 72°C for 2 min) for 10 cycles, (95°C for 15 s, 44°C for 20 s, 72°C for 2 min) for 25 cycles, 72°C for 10 minutes, 4°C for 5 minutes.

Standardly 30 ng of first strand cDNA or 100 ng of gDNA and primers at a final concentration of 1 pm/µl were used in PCR reactions. For PCR Hybaid Gold Taq polymerase (Hybaid, UK) was used at a concentration 1.25 units/ 50 µl. Other reaction concentrations were as follows, 0.2 mM dNTPS (Promega), 1.8 mM $MgCl_2$ (provided with Hybaid Gold Taq polymerases), 1x Hybaid PCR reaction buffer with the volume of reactions made up to 50 µl with ddH_2O. Reactions were set up on ice excluding the Taq polymerase until the reaction mix was had reached 90°C to allow a manual hot start. All PCR reactions were performed on a Hybaid PCRexpress thermal cycler.

PCR products were visualised by running on a 1.4% agarose (SeaKEM agarose, FMC BioProducts, Rockland, USA) modified 1x TAE (40 mM Tris-acetate, pH 8.0, 0.1 mM Na_2EDTA gel with 1 kb ladder markers (Gibco, UK) stained with ethidium bromide to allow product size to assessed. Products were excised using a sterile scalpel and placed in sterile 1.5 ml tubes in preparation for cloning.

6.3.3. Subtractive Screening of inserts

Bacterial colonies containing plasmids with cloned inserts prepared as described below were picked to LB agar plates supplemented with the appropriate antibiotic and grown up overnight. Clones were picked and arrayed in 96 well plate format to all them to be easily screened by PCR. PCR was performed with either the M13R or M13F primer (see below for sequences) and a gene specific primer for the inserts identifies by preliminary sequencing. PCR conditions used were the same as those used for screening for cloned inserts (see below). Subtractive screening on a set of clones was performed with the most common insert identifies by preliminary sequencing an any remaining clones that were not positive for an insert identified by preliminary sequencing were also sequenced (see below)

6.4. 5' and 3' Rapid Amplification of cDNA ends (RACE)

6.4.1 Preparation of RACE ready cDNA

5' and 3' RACE was performed using an adaptation of the method used for the GeneRacer Kit (Invitrogen Life Sciences, USA). 1 ug of total RNA from the species in question was treated with 1 µl calf intestinal phophatase (10 units/ µl) at 50°C with 1 µl RnaseOUT (40 units/ µl) (Invitrogen Life Sciences, USA) made up to a volume of 10 µl with DEPC treated water. This step was designed to remove mRNA transcripts that were not full length

After incubation RNA was precipitated, cleaned and air dried by adding 90 µl of DEPC (Diethyl Pyrocarbonate) treated water and 100 µl of phenol:chloroform, vortexing vigourously for 30 seconds and transferring the aqueous phase to a fresh tube. Next 2 µl of 10 mg/ml mussel glygogen, 10 µl 3 M sodium acetate, pH 5.2 and 220 µl 95% ethanol were added to the aqueous phase and mixed briefly. This mix was then frozen on dry ice for 10 minutes at then spun at maximum speed on a microcentrifuge for 20 minutes at 4°C. The supernatant was then removed and 500 µl of 75% ethanol added to wash the RNA pellet before centrifuging at maximum

speed for a further 2 minutes. Ethanol was then carefully removed and the RNA pellet allowed to air dry for 2 minutes. The pellet was then resuspended in 7 µl of DEPC water for the next stage.

Resuspended RNA was treated with 1 µl (0.5 units/µl) of tabacco acid pyrophosphatase (Invitrogen Life Sciences, USA) in a volume of 10 µl made up with DEPC water at 37°C for one hour. This procedure removes the 5' methyl cap structure of mRNA transcripts. RNA was then cleaned, precipitated and air dried as described above.

Next 0.25 ug of an RNA oligo of known sequence (see Table 3) was then ligated to the 5' end of decapped mRNAs by incubation with 1 µl T4 RNA ligase (5 units/µl), 1 µl 10x Ligase Buffer and 1 µl 10 mM ATP (all supplied by Invirtrogen Life Sciences, USA) in a total volume of 10 µl at 37°C for 1 hr. RNA was then cleaned, precipitated and air dried again.

6.4.2.Reverse transcription of mRNA

Reverse transcription was performed with a dt tailed primer of known sequence (see Table 6.4). RNA was heated to 65°C with 2 pmoles of the dt tailed prime and cilled on ice for 2 minutes. Reverse transcription was performed by adding 1 µl 100 mM dNTPS, 2 µl 10x RT Buffer, 1 µl RNaseOUT and 1 µl AMV-RT (5 units/µl) made up to a volume of 20 µl with sterile water and incubated at 42°C for 1hr. The reaction mix was then heated to 85°C for 15 minutes to deactivate the AMV-RT.

6.4.3 Amplification of cDNA ends

PCR amplification of cDNA ends was used with primers to anneal to the labelled 3' or 5' ends of the cDNA prepared above (see Table 6.4) and either gene specific forward primers (3' RACE) or reverse primers (5' RACE) designed to gene fragments cloned in the degenerate PCR screen (see table 6.4 for sequences of these). Primers were used at a concentration of 1 pm/µl and 2 µl of RACE ready cDNA was

used as template. PCR was performed using Hybaid Gold Taq polymerase (Hybaid, UK) was used at a concentration 1.25 units/ 50 µl. Other reaction concentrations were as follows, 0.2 mM dNTPS (Promega), 1.5 mM $MgCl_2$ (provided with Hybaid Gold Taq polymerases), 1x Hybaid PCR reaction buffer with the volume of reactions made up to 50 µl with ddH_2O. Reactions were set up on ice excluding the Taq polymerase until the reaction mix was had reached 90°C to allow a manual hot start.

Cycling conditions were as follows for both 3' and 5' RACE reactions: 95°C for 1min for 1 cycle, (95°C for 15 s, 68°C to 58°C decreasing by 1°C each cycle for 20 s, 72°C for 2 min) for 10 cycles, (95°C for 15 s, 63°C for 20 s, 72°C for 2 min) for 25 cycles, 72°C for 10 minutes, 4°C for 5 minutes. All PCR reactions were performed on a Hybaid PCRexpress thermal cycler.

PCR products were visualised by running on a 1.2% agarose (SeaKEM agarose, FMC BioProducts, USA) modified 1x TAE gel with 1 kb ladder markers (Gibco, UK) stained with ethidium bromide to allow product size to assessed. Products were excised using a sterile scalpel and placed in sterile 1.5 ml tubes in preparation for cloning. For some reactions where no product was visible 1 µl was of the PCR reaction was used as template in a subsequent using nested PCR primers sets using the same cycling conditions described above. This step was necessary for about half the genes for which some extra flanking sequence was eventually cloned.

Table 6.2. Primers used for Rapid Amplification of cDNA ends

Primer name	Primer sequence
GeneRacer RNA oligo	CGACUGGAGCACGAGGAACACUGACAUGGACUGAAGGAGUAGAAA
GeneRacer 5' Primer	CGACTGGAGCACGAGGACACTGA
GeneRacer nested primer	GGACACTGACATGGACTGAAGGAGTA
GeneRacer dt tail primer	GCTGTCAACGATACGCTACGCTACGTAACGGCATGACAGTG$(T)_{18}$
GeneRacer 3' primer	GCTGTCAACGATACGCTACGTAACG
GeneRacer nested primer	CGCTACGTAACGGCATGACAGTG
Bm-mab-5 F1	AGGCGACAAGAAATTTCTGAAAG
Bm-mab-5 F2	GTGCCGATGAACAGATACCATC
Bm-mab-5 R1	CTTTCAGAAATTTCTTGTCGCCT
Bm-mab-5 R2	GATGGTATCTGTTCATCGGCAC
Bm-egl-5 F1	GAAGAGCTTCGACTACAGACC
Bm-egl-5 F2	GCAGTTATGTGAAAACATATCAACG
Bm-egl-5 R1	GGTCTGTAGTCGAAGCTCTTC
Bm-egl-5 R2	CGTTGATATGTTTTCACATAACTGC
Bm-lin-39 R1	CCAATTCCAATACCTGATTCCTAG
Bm-ceh-13 R1	GTAGGATTGATGCGATTTCGGCTC
Bm-ant-1 R1	AGACCCAACGTTCGGTTTAATTC
Bm-ant-1 R2	GTCGTTCAGTGAGACCCAACG
Bm-hox-3 R1	AGCGCTAATAGTGTTGCAAGTTC
Bm-hox-3 R1	CGTTGCTTATTCAAATAGCGA
Sr-ceh-13 R1	TCACGTCTTCTACTACGATCC
As-ceh-13 R1	TCAATTGCAAAATAGACGCGAT
Ts-hox-3 F1	CAGCAGATATTTGAGCAAGCCG
Ts-hox-3 R1	GGCTTGCTCAAATATCTGCTG
As-hox-3 R1	ACTCAGGTAACGATTCGTCCG
Ts-ant-1 F1	CCGTTATTTGACCAGGAGACG
Ts-ant-1 R1	CTTGACGTTCCGTAAGGCCAAC
Ts-ant-3 F1	CCACTTTAACAGGTATTTGACACG
Ts-ant-3 R1	TCGCTCTGTTAAGCATAATTCGG
Ts-lin-39 F1	CCACTTTAACAGGTATTTGACACG
Ts-php-3 F1	TCCTTTACAATGCTTACGTGTCG

Table 6.2. Primers used for Rapid Amplification of cDNA ends

Gene specific primers and primers provided with the GeneRacer kit were used to amplify were used to amplify Hox gene regions outside that cloned by degenerate PCR. Some amplified products were not full length (see chapter 3). All the gene specific primers described here were involved in RACE reactions that led to the cloning of some extra gene sequence. The sequences of all the Hox genes cloned from nematodes are presented in Appendix 1.

6.5.Inverse PCR protocol

Inverse PCR was not extensively used but led to the identfication of some flanking sequence outside of the region cloned by degenerate PCR for the genes *Bm-ceh-13*, *Bm-lin-39*. The principle of inverse PCR is as follows: you have a small fragment of DNA of known sequence and want to know the sequence of the flanking regions. Genomic DNA is cut with restriction enzymes and ligated in a dilute solution so that the genomic DNA fragments form circular pieces of DNA. Primers are designed within the known DNA sequence to amplify outwards from the known region into the flanking regions of the circularised DNA.

1 µg of *B. malayi* genomic DNA was cut by a restriction enzyme according to the manufacturers protocol using six base cutters that did not cut in the known regions of the *B. malayi* Hox genes. Restricted gDNA was then cleaned by phenol chloroform extraction and precipitated with ethanol as described above. DNA was then resuspended in a volume of 800 µl, including 8 µl (1 unit/µl) T4 ligase, 80 µl T4 ligase 10x reaction buffer and incubated overnight and for the proceeding morning (18 hr) at 16°C.

Ligated DNA was then cleaned and precipitated and quantified using a DNA spectrophotmeter (Genequant, Abgene). Approximately 100 ng was then used in subsequent PCR reactions.

First round PCR reactions were performed using the internal outfacing primer pair and the second round PCR reactions were performed with the external primer pair with 1 or 2 µl of the first round reaction as template. A larger number of different PCR cycling conditions were used in attempt to amplify products from circularised gDNA. The conditions that amplified small 3' extensions for both the *Bm-lin-39* gene and *Bm-ceh-13* gene were as follows for both rounds of PCR: 95°C for 1 min x1 cycle, (95°C for 25 s, 58°C for 20 s, 72°C 2 min) x35 cycles, 72°C for 5 min x1 cycle, 4°C holding step. PCR reactions were performed on a Hybaid PCRexpress thermal cycler.

Reaction components were as follows:1.25 units Hybaid Gold Taq polymerase, 0.2 mM dNTPS (Promega), 2.5 mM $MgCl_2$ (provided with Hybaid Gold

Taq polymerases), 1x Hybaid PCR reaction buffer, 0.08 µl Single strand binding protein (1.5 mg/ml only used in first round reactions, Sigma, UK) with the volume of reactions made up to 50 µl with ddH_2O. Primers were included at a final concentration of 1 pm/µl. Primer sequences used for inverse PCR are presented in Table 6.3.

PCR products were visualised by running on a 1.2% agarose (SeaKEM agarose, FMC BioProducts, USA) modified 1x TAE gel with 1 kb ladder markers (Gibco, UK) stained with ethidium bromide to allow product size to assessed. Products were excised using a sterile scalpel and placed in sterile 1.5 ml tubes in preparation for cloning.

Table 5.4 Inverse PCR primers

Primer name	Primer sequence
Bm-lin-39 inv F1	TGGTTCCAGAATCGTCGTATG
Bm-lin-39 inv F2	CATTCGTTGATGCTAACTGAGCG
Bm-lin 39 inv R1	CTATTTCGATGCGCCTCTTTCGAG
Bm-lin 39 inv R2	GAAATGAAATTCTTTTTCCAATTCC
Bm-ceh-13 inv F1	GAGCCGAAATCGCATCAATCCTAC
Bm-ceh-13 inv F2	TCGAAATATTTAAATCGAACAAGGC
Bm-ceh-13 inv R1	CAAGTTCTGTTAATTGATGAGTGG
Bm-ceh-13 inv R2	GCATTCGTATCCTCTACTACACG

Table 5.4 Inverse PCR primers

Inverse PCR primers were designed to the known sequence of *Bm-lin-39* and *Bm-ceh-13*. The F2 and R2 primers are the internal pairs and the F1 and R1 primers are the external pairs. Nucleic acid sequences of the genes are presented in Appendix 1.

6.6. Cloning of PCR products and screening for positive inserts

PCR products were gel excised and spun through Ultrafree-DA spin columns according to the manufacturers protocol (Millipore, UK). The ellute was placed on 0.0025 μm dialysis membranes (Millipore, UK) over sterile water for 2 hr. The dialysed elute was then concentrated using a DNA vacuum concentrator (DNA SpeedVac, Savant, USA). Concentrated PCR products were resuspended in an appropriate volume for the cloning procedure.

Fragments were cloned using the TOPO TA Cloning Kit for Sequencing (Invitrogen Life Sciences, USA) according to the manufacturers protocols. Ligation reactions were transformed into the competent cells supplied with the kit and plated out on LB agar plates containing 50 μg/ml kanamycin and grown at 37°C overnight. All colonies were assumed to have inserts as any plasmid without an insert is toxic to the competent cells supplied.

Colonies were transferred to LB agar patch plates containing 100 μg/ml ampicillin and grown overnight at 37°C. Bacteria representaive of each colony were then PCR screened using vector primers (M13 Reverse, CAGGAAACAGCTATGA and M13 Forward, GTAAAACGACGGCCAG). A small amount of bacteria were placed in a 20 μl PCR reaction with the following component: 0.5 units of Hybaid Taq polymerase, 0.2 mM dNTPS (Promega), 1.5 mM $MgCl_2$ (provided with Hybaid Gold Taq polymerases), 1x Hybaid PCR reaction buffer with the volume of reactions made up to 20 μl with ddH_2O. PCR conditions were as follows 95°C for 1 min x1 cycle, (95°C for 25 s, 55°C for 20 s, 72°C 3 min) x35 cycles, 72°C for 10 min x1 cycle, 4°C holding step. The extension time at 72°C was increased for particularly large inserts.

3 μl of PCR products were visualised by running on an appropriate percentage agarose (SeaKEM agarose, FMC BioProducts, USA) modified 1x TAE gel with 1 kb ladder markers (Gibco, UK) stained with ethidium bromide to allow

product size to assessed. Some the remaining PCR product could then be used for sequencing if required.

6.7. Sequencing of PCR products and plasmids

Three different types of template were used in sequencing reactions. PCR products from colony screening, primary PCR products and plasmids containing cloned PCR products. Plasmids were made by growing single colonies overnight in 10 ml of LB broth and the appropriate antibiotic and then prepared using the Qiagen plasmid Miniprep kit according to the manufacturers protocol (Qiagen, UK). PCR products from colony screening were cleaned by taking 15µl of PCR product and adding 1µl shrimp alkaline phosphatase (SAP, Gibco, UK) and 1.5 µl exonuclease I (diluted 1:10 in SAP dilution buffer, Gibco UK) and incubating at 37°C for 30 minutes and subsequently at 80°C for 10 minutes. Primary PCR products were cleaned by gel excision and treatment as described above for PCR product cloning.

Sequencing reactions were performed using approximately 150-400 ng of cleaned PCR product depending on length (usually 4 µl of an EXO/SAP treated colony PCR) or 4 µl of a plasmid preparation and either vector primers or gene specific primers in the case of primary PCR products at a final concentration 0.16 pm/µl. The other reaction component was 4µl of ready reaction mix (Perkin Elmer, USA) with the final volume made up to 10 µl using sterile water. Sequencing reactions were performed on a Hybaid PCR express thermal cycler with 25 cycles of 96°C for 30 s, 50°C for 20 s and 60°C 4 min.

Products were then run an ABI 377 automated sequencer (Perkin Elmer, USA) and sequences extracted and checked using Sequence Analysis software (Perkin Elmer, USA). All further sequnce analysis was performed using the software AssemblyLign and MacVector (Oxford Molecular, UK), using a G3 Apple Macintosh Computer.

6.8. RT-PCR to assess transcript levels through the *B. malayi* lifecycle

6.8.1. Obtaining parasite samples for RT-PCR analysis and first strand cDNA synthesis

Embryos were collected from adult female *B. malayi* by dissecting out the posterior and medial regions of gonad on a microscope slide using fine dissection needle. Parasites were overlaid with 150µl of Phospate Buffered Saline (PBS)during dissection. On rupture of the gonad embryos floated to the surface of the PBS which was then gently removed. Parasites were overlayed with PBS a second time and the process repeated. Embryos were collected in a 1.5 ml sterile tube gently spun at 3000 rpm and stored in Trizol (Gibco, UK) at -80°C.

Pre-blood feed microfilaria (Mf) were collected from the peritoneal cavity and blood sacraficed jirds by Ms Yvonne Harcus, to whom I am very grateful. Mf were then cleaned by washing in twice in RPMI 1640 (Gibco, UK) and then gently pelleted in a 1.5ml sterile tube by spinning at 3000 rpm. Mf were then stored in an appropriate volume of Trizol at –80 °C.

Mosquito stage parasites were not collected directly instead 12-14 female *Aedes aegypti* blood fed mosquitoes were collected every day post-blood feed and snap frozen on dry ice in an appropriate volume of Trizol and stored at -80°C.

Total RNA was prepared from all samples, quantified and used to make first strand cDNA as described above with 1µg of total RNA from embryos and preblood fed Mf and 5 µg from each mosquito stage collected (for these 5 separate first strand reactions were performed and then cDNA was concentrated to a volume of 100 µl).

6.8.2. Normalisation of B. malayi transcript levels

Amounts *B. malayi* first strand cDNA (as oppose to mosquito derived) were normalised by assessing the transcript levels of two the house keeping genes *B.malayi* β-tubulin (*Bm-tub-1*) *B. malayi* RNA polymerase II (*Bm-ama-1*). A standardised PCR cycle of 95°C for 1 min x1 cycle, (95°C for 25 s, 55°C for 20 s,

72°C 3 min) x30 cycles was used for both genes. Reaction components were as follows: 1.25 units of Hybaid Taq polymerase, 0.2 mM dNTPS (Promega), 1.5 mM $MgCl_2$ (provided with Hybaid Gold Taq polymerases), 1x Hybaid PCR reaction buffer with the volume of reactions made up to 50 µl with ddH_2O. Primers specific to both transcripts were used at 1pm/µl (see Appendix 2).

The amount of cDNA template added form each lifecycle stage to produce approximately equal amounts of PCR product (quantified by using a DNA spectrophotometer) varied as follows: 1.4 µl from embryonic cDNA, 1 µl from pre-blood feed Mf, 3.5 µl from mosquitoes days 1 and 2, 4µl from mosquitoes days 3-5, 5 µl from mosquito days 6 and 7, 5.5 µl from mosquito days 8 and 9 and 7 µl from mosquito days 10 to 12.

6.8.3. Establishing PCR conditions for each Hox gene

A fragment of approximately 300 bp was amplified from each Hox gene clone using internal primers and sub cloned using the protocols described above. Purified plasmids from these subcloned fragments was then diluted to 1 µg/ml 200 ng/ml and 20 ng/ml for use as PCR templates. PCR conditions that produced high, moderate and small amounts of PCR product were then established empirically using a standard reaction components of: 1µl of diluted plasmid template 1.25 units of Hybaid Taq polymerase, 0.2 mM dNTPS (Promega), 1.5 mM $MgCl_2$ (provided with Hybaid Gold Taq polymerases), 1x Hybaid PCR reaction buffer with the volume of reactions made up to 50 µl with ddH_2O. The PCR conditions established for each gene were slightly different (Table 5.4). Primers were used at 1 pm/µl and all PCRs were performed on a Hybaid PCRexpress thermal cycler.

6.8.4. PCR analysis of transcript level

PCR was performed for each Hox gene with the established conditions (see Table 5.4 and above) using the predetermined ratios of staged cDNA. Reactions were

performed in a volume of 20 µl and 3µl was visualised on a 1.2% agarose gel stained with ethidium bromide. Clear differences were visible between different lifecycle stages (see chapter 3).

Table 5.4 RT-PCR expression analysis PCR conditions

Gene name	PCR cycle conditions
Bm-ceh-13	95°C 30 s x1, (95°C 15 s, 58°C 15 s, 72°C 30 s) x30, 72°C 5 min, 4° hold
Bm-hox-3	95°C 30 s x1, (95°C 15 s, 61°C 15 s, 72°C 30 s) x29, 72°C 5 min, 4° hold
Bm-lin-39	95°C 30 s x1, (95°C 15 s, 62°C 15 s, 72°C 30 s) x29, 72°C 5 min, 4° hold
Bm-mab-5	95°C 30 s x1, (95°C 15 s, 57°C 15 s, 72°C 30 s) x30, 72°C 5 min, 4° hold
Bm-ant-1	95°C 30 s x1, (95°C 15 s, 58°C 15 s, 72°C 30 s) x29, 72°C 5 min, 4° hold
Bm-egl-5	95°C 30 s x1, (95°C 15 s, 60°C 15 s, 72°C 30 s) x31, 72°C 5 min, 4° hold

Table 5.4 RT-PCR expression analysis PCR conditions

PCR conditions to assess the transcript levels of *B. malayi* Hox genes through the parasite lifecycle were optimised empirically by performing PCR on serially diluted plasmid clones.

6.9. Hybridisation to gridded BAC genomic libraries

Genomic clones for some *B. malayi* Hox genes were obtained by PCR to gDNA (see chapter 3) with gene specific primers using the standard PCR conditions outlined above for screening bacterial colonies. The major differences were that annealing temperatures were adjusted to suit the primer pair and for some PCRs a longer extension time was used if no product was generated using a 3 minute extension step. PCR products were cloned as described above accept for products of the Bm-mab-5 gene, one of which could not be successfully cloned (see chapter 3). Partial genomic clones were obtained for *Bm-ceh-13*, *Bm-lin-39*, *Bm-mab-5* and *Bm-egl-5*. These inserts were labelled and hybridised to two high density gridded BAC genomic libraries. As described in chapter 3 only one positive clone was identified for the Bm-egl-5 gene (also containing the Bm-ant-1 homeodomain), no positive clones have been found for the other Hox genes.

Probes were labelled by PCR using biotinylated primers at 1 pm/µl. Reaction conditions and components were as described for colony screening above using

plasmids containing genomic inserts. PCR products were gel excised and cleaned as described above.

Hybridisations were then performed exactly as described in Foster, et al 2001.

6.10. Phylogenetic analysis

All phylogenetic analysis was performed using the alignment presented in Appendix 3.Phylogenetic trees were generated using PAUP 4.08b4. For Neighnour Joining and UPGMA distance analyses total charcter distance was measured with all settings on defaults. For maximum parsimony analysis heuristic searches were performed with MAXTREES set to 10, 000 and with all other settings to default.

The evolutionary rate of Nematode Hox genes was compared to that of arthropods and other groups by comparing orthology groups on a minimum evolution basis. For this purpose a phylogeny representing that presented as the current best estimate of Bilaterian relationships was assumed. Within the Nematoda interrelationships between the species studied here were based on that specified by the analysis of 18S sequnces by Blaxter et al 1998. Within the arthropods the Hexapaoda were allied with the Crustacea as suggested by recent molecular analysis of this phylum(Blaxter, 2001; Giribet et al., 2001). MacClade 4.0 (Maddison and Maddison, 1989) was used to estimate the minimum number of unambiguous changes for each orthology group tested based on the allignment in Appendix 3.

BIBLIOGRAPHY

Aboobaker, A. A., and Blaxter, M. L. (2000). Medical significance of Caenorhabditis elegans. Ann Med *32*, 23-30.

Abramovich, C., Shen, W. F., Pineault, N., Imren, S., Montpetit, B., Largman, C., and Humphries, R. K. (2000). Functional cloning and characterization of a novel nonhomeodomain protein that inhibits the binding of PBX1-HOX complexes to DNA. J Biol Chem *275*, 26172-26177.

Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'homme, B., and de Rosa, R. (2000). The new animal phylogeny: reliability and implications. Proc Natl Acad Sci U S A *97*, 4453-4456.

Aguinaldo, A. M., Turbeville, J. M., Linford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A., and Lake, J. A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature *387*, 489-493.

Akam, M. (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. Cell *57*, 347-349.

Akam, M. (1995). Hox genes and the evolution of diverse body plans. Philos Trans R Soc Lond B Biol Sci *349*, 313-319.

Akam, M. (1998a). Hox genes in arthropod development and evolution. Biol Bull *195*, 373-374.

Akam, M. (1998b). Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. Int J Dev Biol *42*, 445-451.

Akam, M. (1998c). Hox genes: from master genes to micromanagers. Curr Biol *8*, R676-678.

Akam, M. (2000). Arthropods: developmental diversity within a (super) phylum. Proc Natl Acad Sci U S A *97*, 4438-4441.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., *et al.* (1998). Zebrafish hox clusters and vertebrate genome evolution. Science *282*, 1711-1714.

Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H., and Vance, V. B. (1998). A viral suppressor of gene silencing in plants. Proc Natl Acad Sci U S A *95*, 13079-13084.

Aparicio, S., Hawker, K., Cottage, A., Mikawa, Y., Zuo, L., Venkatesh, B., Chen, E., Krumlauf, R., and Brenner, S. (1997). Organization of the Fugu rubripes Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. Nat Genet *16*, 79-83.

Arenas-Mena, C., Martinez, P., Cameron, R. A., and Davidson, E. H. (1998). Expression of the Hox gene complex in the indirect development of a sea urchin. Proc Natl Acad Sci U S A *95*, 13062-13067.

Arthur, W. (2000). The concept of developmental reprogramming and the quest for an inclusive theory of evolutionary mechanisms. Evol Dev *2*, 49-57.

Asahara, H., Dutta, S., Kao, H. Y., Evans, R. M., and Montminy, M. (1999). Pbx-Hox heterodimers recruit coactivator-corepressor complexes in an isoform-specific manner. Mol Cell Biol *19*, 8219-8225.

Averof, M., and Patel, N. H. (1997). Crustacean appendage evolution associated with changes in Hox gene expression. Nature *388*, 682-686.

Balavoine, G. (1997). The early emergence of platyhelminths is contradicted by the agreement between 18S rRNA and Hox genes data. C R Acad Sci III *320*, 83-94.

Balavoine, G., and Telford, M. J. (1995). Identification of planarian homeobox sequences indicates the antiquity of most Hox/homeotic gene subclasses. Proc Natl Acad Sci U S A *92*, 7227-7231.

Bargmann, C. I. (2001). High-throughput reverse genetics: RNAi screens in Caenorhabditis elegans. Genome Biol *2*.

Barstead, R. (2001). Genome-wide RNAi. Curr Opin Chem Biol *5*, 63-66.

Bass, B. L. (2000). Double-stranded RNA as a template for gene silencing. Cell *101*, 235-238.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Billy, E., Brondani, V., Zhang, H., Muller, U., and Filipowicz, W. (2001). Specific interference with gene expression induced by long, double- stranded RNA in mouse embryonal teratocarcinoma cell lines. Proc Natl Acad Sci U S A *98*, 14428-14433.

Bird, D. M., and Riddle, D. L. (1989). Molecular cloning and sequencing of ama-1, the gene encoding the largest subunit of Caenorhabditis elegans RNA polymerase II. Mol Cell Biol *9*, 4119-4130.

Blaxter, M. (1998). Caenorhabditis elegans is a nematode. Science *282*, 2041-2046.

Blaxter, M. (2001). Evolutionary biology. Sum of the arthropod parts. Nature *413*, 121-122.

Blaxter, M., Aslett, M., Guiliano, D., and Daub, J. (1999). Parasitic helminth genomics. Filarial Genome Project. Parasitology *118*, S39-51.

Blaxter, M., and Ivens, A. (1999). Reports from the cutting edge of parasitic genome analysis. Parasitol Today *15*, 430-431.

Blaxter, M., and Liu, L. (1996). Nematode spliced leaders--ubiquity, evolution and utility. Int J Parasitol *26*, 1025-1033.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., *et al.* (1998). A molecular evolutionary framework for the phylum Nematoda [see comments]. Nature *392*, 71-75.

Brooke, N. M., Garcia-Fernandez, J., and Holland, P. W. (1998). The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature *392*, 920-922.

Brown, S. J., Mahaffey, J. P., Lorenzen, M. D., Denell, R. E., and Mahaffey, J. W. (1999). Using RNAi to investigate orthologous homeotic gene function during development of distantly related insects. Evol Dev *1*, 11-15.

Brunschwig, K., Wittmann, C., Schnabel, R., Burglin, T. R., Tobler, H., and Muller, F. (1999). Anterior organization of the Caenorhabditis elegans embryo by the labial-like Hox gene ceh-13. Development *126*, 1537-1546.

Brusca RC and Brusca GJ. 1990. Invertebrates. Sunderland, MA. Sinauer.

Burglin, T. R., and Ruvkun, G. (1993). The Caenorhabditis elegans homeobox gene cluster. Curr Opin Genet Dev *3*, 615-620.

Burglin, T. R., Ruvkun, G., Coulson, A., Hawkins, N. C., McGhee, J. D., Schaller, D., Wittmann, C., Muller, F., and Waterston, R. H. (1991). Nematode homeobox cluster. Nature *351*, 703.

Burglin, T. (1994) in Guidebook to the Homeobox Genes (Duboule, D., ed), pp. 25-71, Oxford University Press, Oxford, England

Carthew, R. W. (2001). Gene silencing by double-stranded RNA. Curr Opin Cell Biol *13*, 244-248.

Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. (2000). Gene silencing in worms and fungi. Nature *404*, 245.

Chan, S. K., Ryoo, H. D., Gould, A., Krumlauf, R., and Mann, R. S. (1997). Switching the in vivo specificity of a minimal Hox-responsive element. Development *124*, 2007-2014.

Chang, C. P., Brocchieri, L., Shen, W. F., Largman, C., and Cleary, M. L. (1996). Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. Mol Cell Biol *16*, 1734-1745.

Chisholm, A. (1991). Control of cell fate in the tail region of C. elegans by the gene egl-5. Development *111*, 921-932.

Clark, S. G., Chisholm, A. D., and Horvitz, H. R. (1993). Control of cell fates in the central body region of C. elegans by the homeobox gene lin-39. Cell *74*, 43-55.

Cogoni, C., and Macino, G. (1999). Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. Nature *399*, 166-169.

Cogoni, C., and Macino, G. (2000). Post-transcriptional gene silencing across kingdoms. Curr Opin Genet Dev *10*, 638-643.

Conraths, F. J., Schutzle, B., Schares, G., Christ, H., Hobom, G., and Zahner, H. (1993). The gene coding for the major sheath protein of Litomosoides carinii microfilariae, gp22, is transcribed in oocytes and embryonic cells. Mol Biochem Parasitol *60*, 111-120.

Cook, C. E., Smith, M. L., Telford, M. J., Bastianello, A., and Akam, M. (2001). Hox genes and the phylogeny of the arthropods. Curr Biol *11*, 759-763.

Cowing, D., and Kenyon, C. (1996). Correct Hox gene expression established independently of position in Caenorhabditis elegans. Nature *382*, 353-356.

Crompton DWT, How much human helminthiasis is there in the world? *J. Parisitology* 1999; 85: 397-403

Cunha, A., Azevedo, R. B., Emmons, S. W., and Leroi, A. M. (1999). Variable cell number in nematodes. Nature *402*, 253.

Davis, G. K., and Patel, N. H. (1999). The origin and evolution of segmentation. Trends Cell Biol *9*, M68-72.

de Rosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., Akam, M., Carroll, S. B., and Balavoine, G. (1999). Hox genes in brachiopods and priapulids and protostome evolution. Nature *399*, 772-776.

Delattre, M., and Felix, M. A. (2001). Microevolutionary studies in nematodes: a beginning. Bioessays *23*, 807-819.

Devaney, E. (1985). Lectin-binding characteristics of Brugia pahangi microfilariae. Trop Med Parasitol *36*, 25-28.

Dick, M. H., and Buss, L. W. (1994a). A PCR-based survey of homeobox genes in Ctenodrilus serratus (Annelida: Polychaeta). Mol Phylogenet Evol *3*, 146-158.

Duboule, D., and Dolle, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. Embo J *8*, 1497-1505.

Eizinger, A., Jungblut, B., and Sommer, R. J. (1999). Evolutionary change in the functional specificity of genes. Trends Genet *15*, 197-202.

Eizinger, A., and Sommer, R. J. (1997a). The homeotic gene lin-39 and the evolution of nematode epidermal cell fates. Science *278*, 452-455.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

Elbashir, S. M., Lendeckel, W., and Tuschl, T. (2001b). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev *15*, 188-200.

Fagard, M., Boutet, S., Morel, J. B., Bellini, C., and Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post- transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc Natl Acad Sci U S A *97*, 11650-11654.

Falciani, F., Hausdorf, B., Schroder, R., Akam, M., Tautz, D., Denell, R., and Brown, S. (1996). Class 3 Hox genes in insects and the origin of zen. Proc Natl Acad Sci U S A *93*, 8479-8484.

Falcone, F. H., Loke, P., Zang, X., MacDonald, A. S., Maizels, R. M., and Allen, J. E. (2001). A Brugia malayi homolog of macrophage migration inhibitory factor

reveals an important link between macrophages and eosinophil recruitment during nematode infection. J Immunol *167*, 5348-5354.

Felix, M. A., De Ley, P., Sommer, R. J., Frisse, L., Nadler, S. A., Thomas, W. K., Vanfleteren, J., and Sternberg, P. W. (2000). Evolution of vulva development in the Cephalobina (Nematoda). Dev Biol *221*, 68-86.

Ferreira, H. B., Zhang, Y., Zhao, C., and Emmons, S. W. (1999). Patterning of Caenorhabditis elegans posterior structures by the Abdominal-B homolog, egl-5. Dev Biol *207*, 215-228.

Ferrier, D. E., and Holland, P. W. (2001). Sipunculan ParaHox genes. Evol Dev *3*, 263-270.

Ferrier, D. E., Minguillon, C., Holland, P. W., and Garcia-Fernandez, J. (2000). The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. Evol Dev *2*, 284-293.

Finnerty, J. R., and Martindale, M. Q. (1998). The evolution of the Hox cluster: insights from outgroups. Curr Opin Genet Dev *8*, 681-687.

Finnerty, J. R., and Martindale, M. Q. (1999). Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. Evol Dev *1*, 16-23.

Fire, A. (1999). RNA-triggered gene silencing. Trends Genet *15*, 358-363.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Foster, J. M., Kamal, I. H., Daub, J., Swan, M. C., Ingram, J. R., Ganatra, M., Ware, J., Guiliano, D., Aboobaker, A., Moran, L., *et al.* (2001). Hybridization to high-

density filter arrays of a Brugia malayi BAC library with biotinylated oligonucleotides and PCR products. Biotechniques *30*, 1216-1218, 1220, 1222 passim.

Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature *408*, 325-330.

Gehring, W. J., Affolter, M., and Burglin, T. (1994). Homeodomain proteins. Annu Rev Biochem *63*, 487-526.

Gellon, G., and McGinnis, W. (1998). Shaping animal body plans in development and evolution by modulation of Hox expression patterns. Bioessays *20*, 116-125.

Giribet, G., Edgecombe, G. D., and Wheeler, W. C. (2001). Arthropod phylogeny based on eight molecular loci and morphology. Nature *413*, 157-161.

Giusti, A. F., Hinman, V. F., Degnan, S. M., Degnan, B. M., and Morse, D. E. (2000). Expression of a Scr/Hox5 gene in the larval central nervous system of the gastropod Haliotis, a non-segmented spiralian lophotrochozoan. Evol Dev *2*, 294-302.

Gomez-Escobar, N., Gregory, W. F., and Maizels, R. M. (2000). Identification of tgh-2, a filarial nematode homolog of Caenorhabditis elegans daf-7 and human transforming growth factor beta, expressed in microfilarial and adult stages of Brugia malayi. Infect Immun *68*, 6402-6410.

Gomez-Escobar, N., Lewis, E., and Maizels, R. M. (1998). A novel member of the transforming growth factor-beta (TGF-beta) superfamily from the filarial nematodes Brugia malayi and B. pahangi. Exp Parasitol *88*, 200-209.

Gonczy, P., Echeverri, G., Oegema, K., Coulson, A., Jones, S. J., Copley, R. R., Duperon, J., Oegema, J., Brehm, M., Cassin, E., *et al.* (2000). Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. Nature *408*, 331-336.

Grandien, K., and Sommer, R. J. (2001). Functional comparison of the nematode Hox gene lin-39 in C. elegans and P. pacificus reveals evolutionary conservation of protein function despite divergence of primary sequences. Genes Dev *15*, 2161-2172.

Greer, J. M., Puetz, J., Thomas, K. R., and Capecchi, M. R. (2000). Maintenance of functional equivalence during paralogous Hox gene evolution. Nature *403*, 661-665.

Gregory, W. F., Atmadja, A. K., Allen, J. E., and Maizels, R. M. (2000). The abundant larval transcript-1 and -2 genes of Brugia malayi encode stage-specific candidate vaccine antigens for filariasis. Infect Immun *68*, 4174-4179.

Grenier, J. K., and Carroll, S. B. (2000). Functional evolution of the Ultrabithorax protein. Proc Natl Acad Sci U S A *97*, 704-709.

Grenier, J. K., Garber, T. L., Warren, R., Whitington, P. M., and Carroll, S. (1997). Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. Curr Biol *7*, 547-553.

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001). Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control C. elegans Developmental Timing. Cell *106*, 23-34.

Grishok, A., Tabara, H., and Mello, C. C. (2000). Genetic requirements for inheritance of RNAi in C. elegans. Science *287*, 2494-2497.

Grosshans, H., and Slack, F. J. (2002). Micro-RNAs: small is plentiful. J Cell Biol *156*, 17-22.

Haase, A., Stern, M., Wachtler, K., and Bicker, G. (2001). A tissue-specific marker of Ecdysozoa. Dev Genes Evol *211*, 428-433.

Halanych, K. M., Bacheller, J. D., Aguinaldo, A. M., Liva, S. M., Hillis, D. M., and Lake, J. A. (1995). Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science *267*, 1641-1643.

Hammond, S. M., Bernstein, E., Beach, D., and Hannon, G. J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., and Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

Harris, J., Honigberg, L., Robinson, N., and Kenyon, C. (1996). Neuronal cell migration in C. elegans: regulation of Hox gene expression and cell position. Development *122*, 3117-3131.

Helm, R., Selkirk, M. E., Bradley, J. E., Burns, R. G., Hamilton, A. J., Croft, S., and Maizels, R. M. (1989). Localization and immunogenicity of tubulin in the filarial nematodes Brugia malayi and B. pahangi. Parasite Immunol *11*, 479-502.

Hintz, M., Hirzmann, J., Hobom, G., Linder, D., Lottspeich, F., Schott, H. H., Conraths, F. J., Zahner, H., and Stirm, S. (1994). Litomosoides carinii microfilarial sheaths: partial amino acid sequences of several major polypeptide constituents. Mol Biochem Parasitol *67*, 69-78.

Hirzmann, J., Schnaufer, A., Hintz, M., Conraths, F., Stirm, S., Zahner, H., and Hobom, G. (1995). Brugia spp. and Litomosoides carinii: identification of a covalently cross-linked microfilarial sheath matrix protein (shp2). Mol Biochem Parasitol *70*, 95-106.

Holland, P. W. (1999). The future of evolutionary developmental biology. Nature *402*, C41-44.

Holland, P. W. (2001). Beyond the Hox: how widespread is homeobox gene clustering? J Anat *199*, 13-23.

Holland, P. W., and Garcia-Fernandez, J. (1996). Hox genes and chordate evolution. Dev Biol *173*, 382-395.

Hope, I. A. (2001). RNAi surges on: application to cultured mammalian cells. Trends Genet *17*, 440.

Hwang, U. W., Friedrich, M., Tautz, D., Park, C. J., and Kim, W. (2001). Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature *413*, 154-157.

Ishii, M., Mitsunaga-Nakatsubo, K., Kitajima, T., Kusunoki, S., Shimada, H., and Akasaka, K. (1999). Hbox1 and Hbox7 are involved in pattern formation in sea urchin embryos. Dev Growth Differ *41*, 241-252.

Jungblut, B., and Sommer, R. J. (1998). The Pristionchus pacificus mab-5 gene is involved in the regulation of ventral epidermal cell fates. Curr Biol *8*, 775-778.

Kennerdell, J. R., and Carthew, R. W. (2000). Heritable gene silencing in Drosophila using double-stranded RNA. Nat Biotechnol *18*, 896-898.

Kenyon, C. (1994). If birds can fly, why can't we? Homeotic genes and evolution. Cell *78*, 175-180.

Kenyon, C. J., Austin, J., Costa, M., Cowing, D. W., Harris, J. M., Honigberg, L., Hunter, C. P., Maloof, J. N., Muller-Immergluck, M. M., Salser, S. J., *et al.* (1997). The dance of the Hox genes: patterning the anteroposterior body axis of Caenorhabditis elegans. Cold Spring Harb Symp Quant Biol *62*, 293-305.

Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., and Plasterk, R. H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.

Ketting, R. F., and Plasterk, R. H. (2000). A genetic link between co-suppression and RNA interference in C. elegans. Nature *404*, 296-298.

Kim JH, 1993. Improving the accuracy of phylogenetic estimation by combining different methods. Systematic Biology 42: 331-340.

Knight, S. W., and Bass, B. L. (2001). A Role for the RNase III Enzyme DCR-1 in RNA Interference and Germ Line Development in C. elegans. Science *2*, 2.

Kobayashi, M., Furuya, H., and Holland, P. W. (1999). Dicyemids are higher animals. Nature *401*, 762.

Kourakis, M. J., and Martindale, M. Q. (2000). Combined-method phylogenetic analysis of Hox and ParaHox genes of the metazoa. J Exp Zool *288*, 175-191.

Kourakis, M. J., and Martindale, M. Q. (2001). Hox gene duplication and deployment in the annelid leech Helobdella. Evol Dev *3*, 145-153.

Kourakis, M. J., Master, V. A., Lokhorst, D. K., Nardelli-Haefliger, D., Wedeen, C. J., Martindale, M. Q., and Shankland, M. (1997). Conserved anterior boundaries of

Hox gene expression in the central nervous system of the leech Helobdella. Dev Biol *190*, 284-300.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191-201.

Kuwabara, P. E., and Coulson, A. (2000a). RNAi--prospects for a general technique for determining gene function. Parasitol Today *16*, 347-349.

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Lavrov, D. V., and Brown, W. M. (2001). Trichinella spiralis mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAS and has a gene arrangement relatable to those of coelomate metazoans. Genetics *157*, 621-637.

Lee, R. C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862-864.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565-570.

Lewis, E. B. (1998). The bithorax complex: the first fifty years. Int J Dev Biol *42(3*, 403-415.

Li, Y. X., Farrell, M. J., Liu, R., Mohanty, N., and Kirby, M. L. (2000). Double-stranded RNA injection produces null phenotypes in zebrafish. Dev Biol *217*, 394-405.

Lipardi, C., Wei, Q., and Paterson, B. M. (2001). RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. Cell *107*, 297-307.

Liu, J., and Fire, A. (2000). Overlapping roles of two Hox genes and the exd ortholog ceh-20 in diversification of the C. elegans postembryonic mesoderm. Development *127*, 5179-5190.

Lohr, U., Yussa, M., and Pick, L. (2001). Drosophila fushi tarazu. a gene on the border of homeotic function. Curr Biol *11*, 1403-1412.

Long, S., and Byrne, M. (2001). Evolution of the echinoderm Hox gene cluster. Evol Dev *3*, 302-311.

Maddison, W. P., and Maddison, D. R. (1989). Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol *53*, 190-202.

Maeda, I., Kohara, Y., Yamamoto, M., and Sugimoto, A. (2001). Large-scale analysis of gene function in Caenorhabditis elegans by high- throughput RNAi. Curr Biol *11*, 171-176.

Maine, E. M. (2000). A conserved mechanism for post-transcriptional gene silencing? Genome Biol *1*.

Maizels, R. M., Blaxter, M. L., and Scott, A. L. (2001). Immunological genomics of Brugia malayi: filarial genes implicated in immune evasion and protective immunity. Parasite Immunol *23*, 327-344.

Maloof, J. N., and Kenyon, C. (1998). The Hox gene lin-39 is required during C. elegans vulval induction to select the outcome of Ras signaling. Development *125*, 181-190.

Maloof, J. N., Whangbo, J., Harris, J. M., Jongeward, G. D., and Kenyon, C. (1999). A Wnt signaling pathway controls hox gene expression and neuroblast migration in C. elegans. Development *126*, 37-49.

Mann, R. S. (1997). Why are Hox genes clustered? Bioessays *19*, 661-664.

Manoury, B., Gregory, W. F., Maizels, R. M., and Watts, C. (2001). Bm-CPI-2, a cystatin homolog secreted by the filarial parasite Brugia malayi, inhibits class II MHC-restricted antigen processing. Curr Biol *11*, 447-451.

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., and Gehring, W. J. (1984a). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell *37*, 403-408.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., and Gehring, W. J. (1984b). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature *308*, 428-433.

McVay, C. S., Klei, T. R., Coleman, S. U., and Bosshardt, S. C. (1990). A comparison of host responses of the Mongolian jird to infections of Brugia malayi and B. pahangi. Am J Trop Med Hyg *43*, 266-273.

Mito, T., and Endo, K. (1997). A PCR survey of hox genes in the sea star, Asterina minor. Mol Phylogenet Evol *8*, 218-224.

Mito, T., and Endo, K. (2000). PCR survey of Hox genes in the crinoid and ophiuroid: evidence for anterior conservation and posterior expansion in the echinoderm Hox gene cluster. Mol Phylogenet Evol *14*, 375-388.

Mouchel-Vielh, E., Rigolot, C., Gibert, J. M., and Deutsch, J. S. (1998). Molecules and the body plan: the Hox genes of Cirripedes (Crustacea). Mol Phylogenet Evol *9*, 382-389.

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Jouette, D., Lacombe, A. M., Nikic, S., Picault, N., *et al.* (2000). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell *101*, 533-542.

Murray, J., Gregory, W. F., Gomez-Escobar, N., Atmadja, A. K., and Maizels, R. M. (2001). Expression and immune recognition of Brugia malayi VAL-1, a homologue of vespid venom allergens and Ancylostoma secreted proteins. Mol Biochem Parasitol *118*, 89-96.

Nielsen C. (1995) Animal Evolution: interrelationships of the living phyla. Oxford, UK: Oxford University press.

Nishikura, K. (2001). A short primer on RNAi: RNA-directed RNA polymerase acts as a key catalyst. Cell *107*, 415-418.

Palauqui, J. C., Elmayan, T., Pollien, J. M., and Vaucheret, H. (1997). Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non- silenced scions. Embo J *16*, 4738-4745.

Palopoli, M. F., and Patel, N. H. (1998). Evolution of the interaction between Hox genes and a downstream target. Curr Biol *8*, 587-590.

Parrish, S., and Fire, A. (2001). Distinct roles for RDE-1 and RDE-4 during RNA interference in Caenorhabditis elegans. Rna *7*, 1397-1402.

Parrish, S., Fleenor, J., Xu, S., Mello, C., and Fire, A. (2000). Functional anatomy of a dsRNA trigger. Differential requirement for the two trigger strands in RNA interference. Mol Cell *6*, 1077-1087.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex [see comments]. Nature *397*, 714-719.

Pendleton, J. W., Nagai, B. K., Murtha, M. T., and Ruddle, F. H. (1993). Expansion of the Hox gene family and the evolution of chordates. Proc Natl Acad Sci U S A *90*, 6300-6304.

Peterson, K. J., Cameron, R. A., and Davidson, E. H. (1997). Set-aside cells in maximal indirect development: evolutionary and developmental significance. Bioessays *19*, 623-631.

Peterson, K. J., and Eernisse, D. J. (2001). Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. Evol Dev *3*, 170-205.

Peterson, K. J., Irvine, S. Q., Cameron, R. A., and Davidson, E. H. (2000). Quantitative assessment of Hox complex expression in the indirect development of the polychaete annelid Chaetopterus sp [see comments]. Proc Natl Acad Sci U S A *97*, 4487-4492.

Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., and Wolberger, C. (1999a). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell *96*, 587-597.

Pollock, D. D., Eisen, J. A., Doggett, N. A., and Cummings, M. P. (2000). A case for evolutionary genomics and the comprehensive examination of sequence biodiversity. Mol Biol Evol *17*, 1776-1788.

Purugganan, M. D., and Suddith, J. I. (1999). Molecular population genetics of floral homeotic loci. Departures from the equilibrium-neutral model at the APETALA3 and PISTILLATA genes of Arabidopsis thaliana. Genetics *151*, 839-848.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

Roch, F., and Akam, M. (2000). Ultrabithorax and the control of cell morphology in Drosophila halteres. Development *127*, 97-107.

Rogers, R., Ellis, D. S., and Denham, D. A. (1976). Studies with Brugia pahangi. 14. Intrauterine development of the microfilaria and a comparison with other filarial species. J Helminthol *50*, 251-257.

Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T., and Pendleton, J. W. (1994a). Evolution of Hox genes. Annu Rev Genet *28*, 423-442.

Ruddle, F. H., Bentley, K. L., Murtha, M. T., and Risch, N. (1994b). Gene loss and gain in the evolution of the vertebrates. Dev Suppl, 155-161.

Ruiz-Trillo, I., Riutort, M., Littlewood, D. T., Herniou, E. A., and Baguna, J. (1999). Acoel flatworms: earliest extant bilaterian Metazoans, not members of Platyhelminthes. Science *283*, 1919-1923.

Ruvkun, G., and Hobert, O. (1998). The taxonomy of developmental control in Caenorhabditis elegans. Science *282*, 2033-2041.

Salser, S. J., and Kenyon, C. (1996). A C. elegans Hox gene switches on, off, on and off again to regulate proliferation, differentiation and morphogenesis. Development *122*, 1651-1661.

Schaller, D., Wittmann, C., Spicher, A., Muller, F., and Tobler, H. (1990). Cloning and analysis of three new homeobox genes from the nematode Caenorhabditis elegans. Nucleic Acids Res *18*, 2033-2036.

Schierwater, B., and Desalle, R. (2001). Current problems with the zootype and the early evolution of Hox genes. J Exp Zool *291*, 169-174.

Schoppmeier, M., and Damen, W. G. (2001). Double-stranded RNA interference in the spider Cupiennius salei: the role of Distal-less is evolutionarily conserved in arthropod appendage formation. Dev Genes Evol *211*, 76-82.

Scott, M. P. (1999). Hox proteins reach out round DNA [news; comment]. Nature *397*, 649, 651.

Selkirk, M. E., Yazdanbakhsh, M., Freedman, D., Blaxter, M. L., Cookson, E., Jenkins, R. E., and Williams, S. A. (1991a). A proline-rich structural protein of the surface sheath of larval Brugia filarial nematode parasites. J Biol Chem *266*, 11002-11008.

Seydoux, G., and Dunn, M. A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. Development *124*, 2191-2201.

Sharp, P. A. (1999). RNAi and double-strand RNA. Genes Dev *13*, 139-141.

Sharp, P. A. (2001). RNA interference--2001. Genes Dev *15*, 485-490.

Shi, H., Djikeng, A., Mark, T., Wirtz, E., Tschudi, C., and Ullu, E. (2000). Genetic interference in Trypanosoma brucei by heritable and inducible double-stranded RNA. Rna *6*, 1069-1076.

Sidow, A., and Thomas, W. K. (1994). A molecular evolutionary framework for eukaryotic model organisms. Curr Biol *4*, 596-603.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K. L., Parrish, S., Timmons, L., Plasterk, R. H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell *107*, 465-476.

Simeone, A., Pannese, M., Acampora, D., D'Esposito, M., and Boncinelli, E. (1988). At least three human homeoboxes on chromosome 12 belong to the same transcription unit. Nucleic Acids Res *16*, 5379-5390.

Smith, H. L. (2000). Investigating development of infective stage larvae of filarial nematodes. Front Biosci *5*, E95-E102.

Sommer, R. J. (1997a). Evolution and development--the nematode vulva as a case study. Bioessays *19*, 225-231.

Sommer, R. J. (1997b). Evolutionary changes of developmental mechanisms in the absence of cell lineage alterations during vulva formation in the Diplogastridae (Nematoda). Development *124*, 243-251.

Sommer, R. J. (1999). Convergence and the interplay of evolution and development. Evol Dev *1*, 8-10.

Sommer, R. J. (2001). As good as they get: cells in nematode vulva development and evolution. Curr Opin Cell Biol *13*, 715-720.

Sommer, R. J., Carta, L. K., and Sternberg, P. W. (1994). The evolution of cell lineage in nematodes. Dev Suppl, 85-95.

Sommer, R. J., Eizinger, A., Lee, K. Z., Jungblut, B., Bubeck, A., and Schlak, I. (1998). The Pristionchus HOX gene Ppa-lin-39 inhibits programmed cell death to

specify the vulva equivalence group and is not required during vulval induction. Development *125*, 3865-3873.

Sommer, R. J., and Sternberg, P. W. (1996). Apoptosis and change of competence limit the size of the vulva equivalence group in Pristionchus pacificus: a genetic analysis. Curr Biol *6*, 52-59.

Stern, D. L. (1998a). A role of Ultrabithorax in morphological differences between Drosophila species. Nature *396*, 463-466.

Sulston, J. E., Schierenberg, E., White, J. G., and Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol *100*, 64-119.

Tabara, H., Grishok, A., and Mello, C. C. (1998). RNAi in C. elegans: soaking in the genome sequence. Science *282*, 430-431.

Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C. C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.

Takahashi, Y., Osumi, N., and Patel, N. H. (2001). Body patterning. Proc Natl Acad Sci U S A *98*, 12338-12339.

Telford, M. J. (2000a). Evidence for the derivation of the Drosophila fushi tarazu gene from a Hox gene orthologous to lophotrochozoan Lox5. Curr Biol *10*, 349-352.

Telford, M. J. (2000c). Turning Hox "signatures" into synapomorphies. Evol Dev *2*, 360-364.

Telford, M. J., and Thomas, R. H. (1998). Of mites and zen: expression studies in a chelicerate arthropod confirm zen is a divergent Hox gene. Dev Genes Evol *208*, 591-594.

Timmons, L., Court, D. L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. Gene *263*, 103-112.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA [letter]. Nature *395*, 854.

True, J. R., and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. Evol Dev *3*, 109-119.

Van Auken, K., Weaver, D. C., Edgar, L. G., and Wood, W. B. (2000). Caenorhabditis elegans embryonic axial patterning requires two recently discovered posterior-group Hox genes. Proc Natl Acad Sci U S A *97*, 4499-4503.

van der Hoeven, F., Sordino, P., Fraudeau, N., Izpisua-Belmonte, J. C., and Duboule, D. (1996). Teleost HoxD and HoxA genes: comparison with tetrapods and functional evolution of the HOXD complex. Mech Dev *54*, 9-21.

Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. Trends Genet *17*, 449-459.

Voinnet, O., Lederer, C., and Baulcombe, D. C. (2000). A viral movement protein prevents spread of the gene silencing signal in Nicotiana benthamiana. Cell *103*, 157-167.

Voinnet, O., Vain, P., Angell, S., and Baulcombe, D. C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. Cell *95*, 177-187.

Voronov, D. A., and Panchin, Y. V. (1998). Cell lineage in marine nematode Enoplus brevis. Development *125*, 143-150.

Voronov, D. A., Panchin, Y. V., and Spiridonov, S. E. (1998). Nematode phylogeny and embryology. Nature *395*, 28.

Wada, H., Garcia-Fernandez, J., and Holland, P. W. (1999). Colinear and segmental expression of amphioxus Hox genes. Dev Biol *213*, 131-141.

Wianny, F., and Zernicka-Goetz, M. (2000). Specific interference with gene function by double-stranded RNA in early mouse development. Nat Cell Biol *2*, 70-75.

Wiellette, E. L., and McGinnis, W. (1999). Hox genes differentially regulate Serrate to generate segment-specific structures. Development *126*, 1985-1995.

Williams, S. A., Lizotte-Waniewski, M. R., Foster, J., Guiliano, D., Daub, J., Scott, A. L., Slatko, B., and Blaxter, M. L. (2000). The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of Brugia malayi. Int J Parasitol *30*, 411-419.

Wittmann, C., Bossinger, O., Goldstein, B., Fleischmann, M., Kohler, R., Brunschwig, K., Tobler, H., and Muller, F. (1997a). The expression of the C. elegans labial-like Hox gene ceh-13 during early embryogenesis relies on cell fate and on anteroposterior cell polarity. Development *124*, 4193-4200.

Wrischnik, L. A., and Kenyon, C. J. (1997). The role of lin-22, a hairy/enhancer of split homolog, in patterning the peripheral nervous system of C. elegans. Development *124*, 2875-2888.

Zamore, P. D. (2001). RNA interference: listening to the sound of silence. Nat Struct Biol *8*, 746-750.

Zang, X., Yazdanbakhsh, M., Jiang, H., Kanost, M. R., and Maizels, R. M. (1999). A novel serpin expressed by blood-borne microfilariae of the parasitic nematode Brugia malayi inhibits human neutrophil serine proteinases. Blood *94*, 1418-1428.

APPENDIX 1 NUCLEIC ACID SEQUENCES OF NEMATODE HOX GENES

B. malayi Hox genes cDNA sequences

Bm-ceh-13

GGTTTAATTACCCAAGTTTGAGGCTAGTGAATTCATTTCTCGATGAGTTGTACCGAATCTTACATTGG
TTCTTCTAGTCCACCCACGACTTATCCACATTATGCTGACTGGTCATCGTCCGGTGGTTACATTTCCA
CCTCGGTAACACCTTATCCACCAACCTCATCATATGGTCATCAATTACCTCCACCAAGCACTTCACAT
TGGATTACAAATGCTACAGCTAATTATGGTCAAAATCAATCGACAACTAATGCTAATCCTATCCAACA
ATCTCATTCAACTACTTCTACCACTGCTACTACTGCTGCTACGGTTGCAGCTGCAGCTGCAGCTGCTA
CAATTGCTGCAAGCACCGCAAATACATATAAATGGATGCATATTAAACGTACCGTCACTAAATCATCA
GTACCAAAACGACGTGTAGTAGAGGATACGAATGCGATACGTACAAATTTTACCACTCATCAATTAAC
AGAACTTGAAAAAGAATATTACACATCGAAATATTTAAATCGAACAAGGCGAGCCGAAATCGCATCAA
TCCTACAACTAAATGAAACTCAAGTAAAGATATGGTTCCAAAATCGTCGT

Bm-hox-3

GGTTTAATTACCCAAGTTTGAGGTACTTATGAATCAAACAAATATTATCAGCACAGACACCTCTAATG
ATTATCATCAAACTCAAGCATACTATCCACATTATCGTGCATCAATTGCATCTACCTCTGATTATTCA
CAGTTATTTCTAAATGGTACCATTCCAAATAACTACTCTTCTGCCACTACTACTACTGCTGCTGCTAC
CACAGCAAATTATAGTAGCATTAATTATGGTAGAGGAAATCGAAGTAATCATCAAAGACCTAAATATG
CGTGGATGCTAGAACGTGACAAGGATCATCACCAATCACGTCATTTACAGCAAACTGTTAGATCTACA
ACTGATACCGTAATCACATCACAAACAGGTCGTACATCATATAGTACACCACAGGTTGTTGAATTGGA
AAAAGAATTCCGTACCAATCGCTATTTGAATAAGCAACGACGAAATGAACTTGCAACACTATTAGCGC
TTACTGATCGACAAATTAAAATTTGGTTTCAAAATAGACGAATGAAAGAGAAGAAGCAACGATTAGCT
GCTGCTCTACCACATAAACTTATTGTTACTACTAGTACTGCTACTCCTGCTACCCTTCTTAAAGATAA
ACAAACTCTGCGTTGTTACCACTGCTT

Bm-lin-39

GGTTTAATTACCCAAGTTTGAGTGAATGATTAATATCGAAAGATATGACATCAGCAGAGCAATTTGCA
TCAGAGATCTCAAATTATTACTATGATCAACAACAGTCAAATCCGAGACTTACAACCACCAATAATTC
CGCCGCAGTACCGACATCACTCTCCGCTCCAACATATTATCCAAATGCGCAATACAGTGCTCCTGATT
CCAGTCAACATGGTCCGAGCTCATCATGTAGTTACAATAATGGAGAATGGAATAAAAATGGAAAAGAT
GTTGAAAAAGAGGAGGATGATGAAGAGAAGAATAATGTGGCCGTATATCCATGGATGACACGAGTTCA
TTCAGCTAGTGGTTCAAATCGAGGTGAAAAACGTCAACGGACAGCGTATACTAGGAATCAGGTATTGG
AATTGGAAAAAGAATTTCATTTCAACAAATATTTAACTCGAAAGAGGCGCATCGAAATAGCTCATTCG
TTGATGCTAACTGAGCGACAGGTCAAAATTTGGTTCCAGAATCGTCGTATGAAGCCHAAAAAAAAAAT
AAA

Bm-mab-5

GGTTTAATTACCCAAGTTTGAGTTCACGTCGATGTATCCCAGTTGGCAAAGTCCTGATGATTCGCCTA
CCGTACAGCCTTACTGGAGTCAACACGTTAGTCAGGCAGCTAGTGCCGCAGCTACGCATCAAAAAGTT
GCGGTTGCAGCACCAGCTTATGATCCGCTGCTATATCCTGCCCAAAATTATATGAATAATATGAAAAA
TATGCTTGGTGCCGCTGCGCAATGGGTTGAATCATCAAGCACATTTCCAAACTATAATGCTTTACAAC
CAGCGTCAACATCGCATTTGCCGGTGATAAGTGGTGAACGAGCGGTGCAAATTAACCAGCCGGTGTTT
CCATGGATGAAAATGAGTGGTGGTAAAGGTGGTGAATCAAAAAGGACGCGACAAACATATAGTAGAAA
TCAGACGTTGGAACTTGAAAAAGAATTCCACTATAATAAATATTTAACGCGAAAAAGGCGACAAGAAA
TTTCTGAAAGTTTACAATTATCTGAGCGGCAGGTGAAAATATGGTTCCAGAATAGGCGAATGAAGCAT
AAAAAAGAATGTAAAGGTGAAGTAACAACAGTAAATGAAAGTGATGAAAGTGCCGATGAACAGATACC
ATCAACATGACAAGGATTAGTGGGATAAAGGTTTTAAATAAAAAAAAAAAAAAAAA

Bm-ant-1

GGTTTAATTACCCAAGTTTGAGCGATGGGTTTCGAATTTCCGGCTTCCGGTTCACGAACTTTTCAGGA
TGCGGCAGCAACAGCTGCAGTAGCTGCTGTTATTCCAACAGATGATCATCTACAACGGCTTGCACAAA
TGACTCAAGGTGTTGGTCATACTAATTCATCGTCAACCACTACAAAATCTTCTGGTTCCTCCGATAGC
AGTGACCCAAATTCAACTACCACTCAACCGGATCCAACCTTATATCAGAGCGCTGCTGCTGCATATGT
ACAGCAATATGGTAATTGGCCTAATTATTATCAACAATTTGGACAGCCGATGAATCCAGCTGCGGCTG
CATTTTCTGCCTGGCCAGCACAATGTTATGCACCTCCACATTGGCCAAATTATGATAATTGTAAGAAG
AGAGGAAGACAAACGTATAGTCGTCAGCAAACGCTGGAACTTGAGAAGGAATTTCACTACAACAAATA
TTTGACCCGACGGCGGCGCATTGAATTAAACCGAACGTTGGGTCTCACTGAACGACAAATCAAAATTT
GGTTTCAAAATCGACGGATGAAGAAAAAAAAAGAGGATAAGGCAAGGGATGAGCACAGGAATATTCAT
AATTTTGGAAAAAA

Bm-egl-5

GGTTTAATTACCCAAGTTTGAGCGATGGGTTTCGAATTTCCGGCTTCCGGTTCACGAACTTTTCAGGC
GGCAGCAACAGCTGCAGTAGCTGCTGTTATTCCAACAGATGATCATCTACAACGGCTTGCACAAATGA
CTCAAGGTGTTGGTCATACTAATTCATCGTCAACCACTACAAAATCTTCTGGTTCCTCCGATAGCAGT
GACCCAAATTCAACTACCACTCAACCGGATCCAACCTTATATCAGAGCGCTGCTGCTGCATATGTACA
GCAATATGGTAATTGGCCTAATTATTATCAACAATTTGGACAGCCGATGAATCCAGCTGCGGCTGCAT
TTTCTGCCTGGCCAGCACAATGTTATGCACCTCCACATTGGCCAAATTATGCACAATCGAAGAAGGGT
CGTCAAACATATCAACGTTATCAGACATCGGTACTCGAAAGCAAATTCCAGCAAAGCAGTTATGTGAA
AACATATCAACGTGAAGAGCTTCGACTACAGACCAATTTGAGTGACCGACAAATCAAAATATGGTTTC
AAAATCGTCGAATGAAAGCCAAGAAGGAGAAGAATCGATGCGAGGAAAAATGAGGAAAGTTTGGAATC
TGAAGAAATAAACGACAAAAGCTATAGTTCTACTAATACTGCGGGAAATCAAACTAAAATCGAAAAGC
CAACAGGCGCCAAAAAAAA

Bm-php-3

GGTTTAATTACCCAAGTTTGAGGTACTTATGAATCAAACAAATATTATCAGCACAGACACCTCTAATG
ATTATCATCAAACTCAAGCATACTATCCACATTATCGTGCATCAATTGCATCTACCTCTGATTATTCA
CAGTTATTTCTAAATGGTACCATTCCAAATAACTACTCTTCTGCCACTACTACTACTGCTGCTGCTAC
CACAGCAAATTATAGTAGCATTAATTATGGTAGAGGAAATCGAAGTAATCATCAAAGACCTAAATATG
CGTGGATGCTAGAACGTGACAAGGATCATCACCAATCACGTCATTTACAGCAAACTGTTAGATCTACA
ACTGATACCGTAATCACATCACAAACAGGTCGTACATCATATAGTACACCACAGGTTGTTGAATTGGA
AAAAGAATTCCGTACCAATCGCTATTTGAATAAGCAACGACGAAATGAACTTGCAACACTATTAGCGC
TTACTGATCGACAAATTAAAATTTGGTTTCAAAATAGACGAATGAAAGAGAAGAAGCAACGATTAGCT
GCTGCTCTACCACATAAACTTATTGTTACTACTAGTACTGCTACTCCTGCTACCCTTCTTAAAGATAA
ACAAACTCTGCGTTGTTACCACTGCTT

P.pacificus **Hox genes**

Pp-ceh-13

GAGCTTGAGAAGGAGTTCCATACAAATAAATACGTGAATCGACAAAGGCGGACGGACATTGCTGCCCA
ATTGAAACTGAATGAGGCACAGGTGAGGATCAGTTGGTACTCTTCGTGATATGTATCATCTTGGTTAC
ATGTATAATGTGGGCTTGATGTAGGATGTAATCAGTGACATCAGCGTTGTTTACTCTTTTCATGGGGT
CGTTGAGGAACCAGCTAATTTGGTGTTCCTCATTTCCACCAATCAAATACCTCGTTCCTAATTTCCTC
ACTATCGCTAATTAGACTACGGTAGCTACCGAATTAACTTGAACAGCCTTAATTGCTCTAATTTCAGG
TGAAAATCT

This sequence includes an intron in the homeodomain.

Pp-egl-5

CAGACGTCCGTTCTCGAACAGAAATTCCTCCAATCCTCGTATGTCTCCAAGAAGCAACGGGAAGAATT
ACGAATGCAGACGAATCTCACAGATCGGCAGATTAAAATA

Pp-php-3

CTGCAGACTCTTGAATTGGAGAAGGAATTCCTGTATAATCCCTACGTTTCCAAGCAGAAACGATATGA
ATTAGCCATCAACTTAGGATTGACTGAACGACAAGTCAAAATT

T. spiralis Hox genes

Ts-ceh-13

CTTGTTGAACAGTGGATGTACATCCGTGCCCAACCGCACCAACTTCACCACGAAGCAGTTGACAGAAT
TGGAGAAAGAGTTTCACACGAATCGTTATTTGACCCGTGCCCGTCGTATTGAAATCGCATCCCAGCTC
GGCTTGAACGAGACCCAGGTGAAGATA

Ts-hox-3

AAACGACAACGAACAGCGTACACGAATCGCCAGTTGGTCGAATTGGAAAAAGAATTCCACTTCAGCAG
ATATTTGAGCAAGCCGAGACGTCAAGAATTGGCCGAATCGTTGTCGCTGAGCGAGAGGCAAATTAAAA
TTTGGTTCCAAAATCGACGGATGAAGATGAAAAAAGACGAACGAAGTCGCCGATTTCGCATCACGTTT
TGTTTGAACTGTTGTCTCGGACAAAATTTACAGTAATAAAAACTAGGACGAGCAAAAAAAAAA

Ts-lin-39

CTAGAATTTGAGAAGGAGTTTCATTTCAACCGCTATTTGACCCGTCGTCGACGTGTTGAAATTGCGCA
CACTTTAGATTTAACAGAACGTCAAGTTAAAATTTGGTTCCAAAATCGCCGCATGAAGCTCAAAAAAG
ATCAAAAAACTTTACAACCATCTGCGTTCAGCTCTGCTGCCATAAGCAGCATCATCAATCCGTTGCAG
ATGAACCATCTAACCGCGGCAGCAATGCATGGAACGAGGCCAGGCGCCAGTTATTTGTTCAACGTACC
CTATGACATGCAAGGAAGTTTAAACTCCAATTGCCCTGGTGCGCAAGGTCCAAAAATTTTTTTACAAG
ACTGAAGAACCAATTCTAAATATTCCTCCTCTCAGTTCCATCTTCTGTTTTGCAAACTCTTTTCTTAT
AATTTGCTATTCTGAATATCTCTGAAATTGATCTCTTTCTGATTTCGTTTTTCTCTTTTGGTTGCTTT
GTGTAGCTTACATCAGCCAGTTGTTTACTAGCTTTAATTGAGCAAATTTCGAAGCTGAACTATGGCAA

Ts-ant-1

GCATGCTTGATTTGCCAAGGTAAGCAGCTTGTTTTTTGTTTAGTTGTTACCGGATCTGAGGGTCCCAG
AAAACGAGGTCGACAAACGTATCATCGTTCGCAGACGCTCGAGTTGGAGAAGGAATTTTTCACCAACC
GTTATTTGACCAGGAGACGTCGGATTGAATTAGCCCAGTATGTTGGCCTTACGGAACGTCAAGTGAAA
ATATGGTTCCAAAATAGGCGCATGAAGTGGAAAAAGGAACACAAGCAAAAAGAGCCGGATGATCAGAA
ATCGAATGAAAGCCTTGACGGTTCGAATGTGTCGAGTATCAGCAACTGTAGCTCTACAACGCTTGTCA
ACACAAATCCGGCCACTCCTGCGTCATCACAGCCGTCGATCTCTAAACTGGAATCCAACCAGCGTCAG
AACTATTCAGTGACGGTAGAAGAAACGAAAGATGTCAAAAAAGTCATGTTTGGCCGATGCTGAACAGT
CGTTATTAAAAATTTTCAAGTTGAAAA

Ts-ant-2

CTAGAGTTGGAGAAGGAGTTTCACTACATCAAATATTTGACCCTACCACGGCGCATTGAATTAACCCG
AACGTTGGGTCTAACTGAC

Ts-ant-3

AGATCCAGTTGTTTCACTTGCAAATTGATTTTTCACGTAATCATGCATTCTGCATATGCTTTCCCCGC
TATTGACAGAGCAATAAACCAGACGAGCAATATACATATTATTATCCGTAAGGGCGGTGATCAAAAGC
GAACTCGTCAGACTTATTCGCGAAATCAGACACTCGAATTGGAAAAAGAGTTCCACTTTAACAGGTAT
TTGACACGTCGTCGACGTCAAGAAATTGCTGCCGAATTATGCTTAACAGAGCGACAGGTGAAAATATG
GTTTCAGAATCGTCGAATGAAATGGAAGAAAGAAAATAAAATTTCCGTCGAGCCCGCTGATTCAAGCA
CTGAAAAAGATTTATTCGAATGACGATGTGCTTGTTTTCCAGAATAGTAGTTTATCAAAAAAAAACAA
AACAAACTACTACTACTGTACGTTGTTCATTAGAGCAAACTGAATAAGCAAATGCTGCAGCTCAGTTC
AACTTTTCGATGAAATCTTCTTATGCATGCGATAGTTTATTTATTATTTAAAAAAAAGTTCTAGAACA
TACAAAGATGAAATGTATCCTTCATTTACTTAATGCATGACTTTTGAGGATCTTTTTTTTCAAATCTG
TTTTTCGATATGTCAACTAAAAGGAATAAATATTGCCGTGCATGAAAAAAA

Ts-egl-5

ATCAGCACATCGGTACTCGAAAGCAAATTCCAGCAAAGCAGTTATGTGAGTAAAAAGCAACGTGAAGA
GCTTCGACTACAGACCAATTTGAGTGACCGACAAATCAAAATTA

Ts-php-3

AAGGAGTAGAAAATTTTTTTTCTGAAGATCACAGTATTGAGTGGACGAATGCACTGTCAATACGGAAA
AAGCGGAAGCCGTATACAAAGTACCAGACGTTGGAGTTGGAAAAAGAATTCCTTTACAATGCTTACGT
GTCGAAGCAGAAACGTTGGGAACTCGCCCGAAGTTTGGGTCTTACGGAACGACAAGTGAAAATTTGGT
TTCAAAATCGTCGGATGAAGCAAAAAAAGCAGCAGCAGCGTGCTCAAGCCGAAGCGACAAGCACATCC
CAAGCCCAACATGCACTTGGCTCTTCGTCTACTACTCCGACGTGAAACAGAAGCTGATTTCCAGATGA
TGGCTATTTCGTTTTTTGAACGAAGCTAGCTGATGATGGAATTATAAAATGAGCAAAAATGTTTGCTC
TCTCAATTGCGCGTTGACATCGCCTTTGTTTTTTTTCTTTCGTTATTCTTTCTTTTCATTTTGTATAA
ATTAACATCAACAATGCTCTCTCATTAATTTTCTTATTATTCATTATGGATTAATAAGAGAAAAAAGG
ACATAAA

***A. suum* Hox genes**

As-ceh-13

CGAACAAATTTTACTACTCATCAATTAACCGAACTCGAAAAAGAGTACTACACATCAAAATATTTGAA
TAGAACACGACGAGCTGAAATCGCGTCTATTTTGCAATTGAATGAAACTCAGGTAAAAATC

As-hox-3

GGTCGAACAGCCTATAGCACTCCACAAATTGTCGAATTGGAAAAAGAATTCCGGACGAATCGTTACCT
GAGTAAAGTGCGTCGCAGTGAATTGGCTGAATTATTATCGTTGTCGGATCGCCAAATCAAAATT

As-lin-39

CTAGAGTTGGAGAAGGAATTCCACTTCAACAAATATCTTACACGGAAGCGACGTATCGAAATTGCACA
TTCTCTTATGCTAACCGAACGACAGGTGAAGATT

As-mab-5

CTAGAGTTTGAGAAAGAGTTCCACTATAATAAGTATTTGACAAGGAAGCGAAGGCAGGAGATCTCGGA
GAGTCTTCAGCTGTCCGAAAGACAGGTAAAGATAA

As-ant-1

AGTAGACATCAGACACTGGAGCTGGAAAAGGAGTTCCATTACAACAAGTACCTGACACGTCGTCGTCG
AATCGAGCTTAATCGCAGTTTGGGTCTCTCGGAACGTCAGATCAAGATT

As-egl-5

TCGGTGCTCGAAAACAAGTTCCAGCAGAGCAGTTACGTTAGCAAAAAGCAACGCGAAGAGCTGCGTCT
TCAAACGAATCTCACCGATCGACAGATCAAAATCTGGTTCCAAAACGCGGCTATG

As-php-3

CTAGAGTTTGAGAAGGAGTTCCTGTATAATCCCTACGTTTCCAAGCAGAAACAATATGAATTAGCCAT
CAACTTAGGATTGACTGAACGACAAGTCAAGATT

***S. ratti* Hox genes**

Sr-ceh-13

CGAACAAATTTTACTACGCACCAATTAACAGAATTGGAAAAAGAATATTACACTTCAAAGTATTTGGA
TCGTAGTAGAAGACGTGAAATAGCTAAACAATTGGCATTAAATGAGACACAAGTAAAGATA

Sr-lin-39

AGACAACGAACGGCATACACACGTTCACAGGTATTGGAACTAGAAAAAGAATTTCACTTTAATAAATA
TTTAACTCGTAAACGTCGTATTGAAATAGCACATTCATTAATGTTAACAGAACGACAAGTCAAGATA

Sr-mab-5

CCTGGAGCTGGAGAAGGAATTTCATTTTAACAAATATTTAACGAGAAAAAGACGACAAGAGATCTCAG
AAGCTTTACAATTGACTGAAAGACAGGTGAACTAAAGATAACTTTTTATTATTCTAATTATTTTATTT
AGGTCAAAATT

This sequence includes a small intron in the homeodomain.

Sr-ant-1

TATCAGACTCTCGAACTCGAAAAAGAATTCAAATTCAACCGTTACCTCACACGGCGCAGGCGCGTGGA
GCTATCGGCACTACTCTGCCTCAGTGAACGTCAAATCAAAATA

Sr-egl-5

AGGTATCAGACATCAGTATTGGAATCTAAATTCCAACAAAGTTCTTATGTTTCAAAAAAACAACGTGA
AGAATTGAGACTTCAAACAAATTTAACAGATAGGCAGATAAA

Sr-php-3

AAACGTAAGCCTTATACAAAATTTCAAATATTAGAATTGGAAAAAGAATTTGCCTACAGTAGTTATGT
TAATAAACAGAAACGCGTAGATTTGGCACAACAGTTACAGTTAACAGAAAGACAAGTCAAAATA

M. javanica **Hox genes**

Mj-ceh-13

CTAGAGCTGGAGAAGGAGTTTCATACAAACAAATATCTAAACAAAAGTAGAAGGGCAGAAATTGCCGC
AATGTTACAGCTACACGAATCTCAAATTAAAATA

Mj-hox-3

CGGACGAATCGTTACCTGAGTAAAGTGCGTCGCAATGAATTGGCTGAATTATTATCGTTG
CCGGATCGCCAAATCAAAATAT

Mj-lin-39/1

ACACTGCTTACACACGAAATCAAGTATTAGAGCTTGAAAAAGAATTCCATTTTAACAAGTAAGAGATA
GTTCTTTGTAATACTGATTAAACTAATTTAGATATTTGACAGGAAAAAGGAGGATTGAGATAGCCCAT
ACACTTATACTTACAGAGAGACAGGCATGTATTTATAATTTTTGAAAGATATTCAAATTTTCTTTCAG
GTAAAGATATGG

This sequence includes two small introns in the homeodomain.

Mj-lin-39/2

CTAGAACTTGAGAAGGAATTCCATTTTAACAAATATCTGACAAGAAAAAGGAGGATTGAGATAGCCCA
TACACTTATACTTACAGAGAGACAGGCATGCATTTATATTTTTTGGAGGATATTCAAATTTTCTTTCA
GGTGAAGATATGG

This sequence includes one small intron in the homeodomain.

Mj-mab-5

CGGATCCCTGGAGCTGGAGAAGGAATTTCATTTTAACAAATATTTAACGAGAAAAAGACGACAAGAGA
TCTCAGAAGCTTTACAATTGACTGAAAGACAGGTGAACTAAAGATAACTTTTTATTATTCTAATTATT
TTATTTAGGTCAAAATTTGGTTCCAAAACGCGGCTATG

This sequence includes a small intron in the homeodomain.

APPPENDIX 2 SEQUENCES OF THE *BM-AMA-1*, *BM-TUB-1* AND *BM-SHP-1* AND CONSTRUCTS USED FOR RNAI IN *B. MALAYI*

Bm-ama-1

```
    5   10   15   20   25   30   35   40   45   50   55   60   65   70
CTGAATACTTTTCATTACGCTGGCGTTTCGGGCAAAAATGTCACACTTGGCGTACCGCGGCTTAAAGAAA
 L  N  T  F  H  Y  A  G  V  S  A  K  N  V  T  L  G  V  P  R  L  K  E>

   75   80   85   90   95  100  105  110  115  120  125  130  135  140
TCATCAATGTTTCGAAGAAGCCGAAAACACCTTCATTAACAGTTTTTCTGCAGGGAACTGCTGCGAAGTT
I  I  N  V  S  K  K  P  K  T  P  S  L  T  V  F  L  Q  G  T  A  A  K>

  145  150  155  160  165  170  175  180  185  190  195  200  205  210
TGTCGTTTTTATCCGAAACTTTGTAAAAAATTTTTATTTGAACAAGTGGCGATATTACTTGGTTATTGGA

  215  220  225  230  235  240  245  250  255  260  265  270  275  280
TGATCTATCTTGTTTAAAAAGGACAAATAAATATAACTTGATGAATAGAAGATAAAATTAGTTTTAAATT

  285  290  295  300  305  310  315  320  325  330  335  340  345  350
TCAGAGATGCTGAAAAGGCTAAGGACGTACTGTGTAAGTTGGAACATACAACATTACGAAAGGTAACTGC
      D  A  E  K  A  K  D  V  L  C  K  L  E  H  T  T  L  R  K  V  T  A>

  355  360  365  370  375  380  385  390  395  400  405  410  415  420
AAATACAGCAATTTATTATGATCCAGACCCGAAGAATACCATAATTGAAGAAGATCAGGAATGGGTAAAC
  N  T  A  I  Y  Y  D  P  D  P  K  N  T  I  I  E  E  D  Q  E  W  V  N>

  425  430  435  440  445  450  455  460  465  470  475  480  485  490
ATTTTCTATGAAATGCCGGATTTTGATCCATCTCGATGCAGTCCATGGCTTCTGCGTATCGAATTGGATC
  I  F  Y  E  M  P  D  F  D  P  S  R  C  S  P  W  L  L  R  I  E  L  D>

  495  500  505  510  515  520  525  530  535  540  545  550  555  560
GTCGACGCATGACTGATAAGAAATTAACGATGGAAGCAATTGCAGATAAGATACATCAGGGATTTGGTGA
R  R  R  M  T  D  K  K  L  T  M  E  A  I  A  D  K  I  H  Q  G  F  G  D>

  565  570  575  580  585  590  595  600  605  610  615  620  625  630
TGATCTGAATGTCATTTATACAGATGATAATGCAGAAAAACTTGTATTTCGTTTGCGTATCACCAATCAG
  D  L  N  V  I  Y  T  D  D  N  A  E  K  L  V  F  R  L  R  I  T  N  Q>

  635  640  645  650  655  660  665  670  675  680  685  690  695  700
GAAGGTGACAAGGGTAATGAAGATGAACAAGTGGAGCGAATGGAAGATGACGTGTTCTTGCGTTGTATTG
  E  G  D  K  G  N  E  D  E  Q  V  E  R  M  E  D  D  V  F  L  R  C  I>

  705  710  715  720  725  730  735  740  745  750  755  760  765  770
AAACCAATATGCTCAGTGACCTTACGCTTCAAGGTATTGAAGCTATTACCAAAGTATACATGCATAAACC
E  T  N  M  L  S  D  L  T  L  Q  G  I  E  A  I  T  K  V  Y  M  H  K  P>

  775  780  785  790  795  800  805  810  815  820  825  830  835  840
AACAACTGATGACAAAAAACGAGTTGTGATTACTCCGGATGGTGGTTTTAAGGCAATTCCAGAATGGCTT
  T  T  D  D  K  K  R  V  V  I  T  P  D  G  G  F  K  A  I  P  E  W  L>

  845  850  855  860  865  870  875  880  885  890  895  900  905  910
CTAGAAACTGATGGTACTGCCTTAGCTAAGGTACTGAGTGAACAAAATGTGGACCCGATACGTACGACGT
  L  E  T  D  G  T  A  L  A  K  V  L  S  E  Q  N  V  D  P  I  R  T  T>

  915  920  925  930  935  940  945  950  955  960  965  970  975  980
CCAACGATATATGTGAAATATTCGAAGTTTTGGGAATTGAAGCAGTTCGGAAAGCTATTGAACGTGAAAT
S  N  D  I  C  E  I  F  E  V  L  G  I  E  A  V  R  K  A  I  E  R  E  M>

  985  990  995
GAATCATGTCATTTG
  N  H  V  I
```

This fragment of Bm-ama-1 was cloned by degenerate PCR using primers designed to alignment of RNA polymerase II sequences. The targeting construct was made form a 355bp fragment amplified from this sequence (primers in turquoise). Another primer set was used to assess transcript levels in single worm RT-PCRs (primers in green) This primer set spanned an intron present in the genomic copy of the gene (in yellow) and were also used to normalise transcript levels across cDNA from different lifecycle stages (see chapter 3).

Bm-shp-1

```
    5   10   15   20   25   30   35   40   45   50   55   60   65   70
GGTTTAATTACCCAAGTTTGAGTTATGTGTTGTAAGCTTATTCCTTCATTTTGCATGCTTTTATCGGTAG
                        M  C  C  K  L  I  P  S  F  C  M  L  L  S  V>
   75   80   85   90   95  100  105  110  115  120  125  130  135  140
CAAATGCAGTGCATTTGCATTATGGTCCGCAAGAAGAACATTATCAGGATGTGCAACCAGGACAAGAAAT
A  N  A  V  H  L  H  Y  G  P  Q  E  E  H  Y  Q  D  V  Q  P  G  Q  E  M>
  145  150  155  160  165  170  175  180  185  190  195  200  205  210
GCAACGGAAAAGTATAAAATTACTACGGATGGGACCGCAGATTATGCAACCGCAAGGTATGCTGCCACAA
  Q  R  K  S  I  K  L  L  R  M  G  P  Q  I  M  Q  P  Q  G  M  L  P  Q>
  215  220  225  230  235  240  245  250  255  260  265  270  275  280
GGTATGCAACCACAAGGTATGCAACCACAAGGTATGCAACCACAAGGTATGCAACCACAAGGCATGCAAC
 G  M  Q  P  Q  G  M  Q  P  Q  G  M  Q  P  Q  G  M  Q  P  Q  G  M  Q>
  285  290  295  300  305  310  315  320  325  330  335  340  345  350
CACAAGGCATGCAACCACAAGGTATGCAACCGCAAGGTATGCGACCACAAGGTATGCAACCACAAGGTAT
P  Q  G  M  Q  P  Q  G  M  Q  P  Q  G  M  R  P  Q  G  M  Q  P  Q  G  M>
  355  360  365  370  375  380  385  390  395  400  405  410  415  420
GCAACCGCATTACATCGAACCACAAGCTGTTGATAAAAAGTGTGCTGGGTGCATAATAAATATAAATTGT
  Q  P  H  Y  I  E  P  Q  A  V  D  K  K  C  A  G  C  I  I  N  I  N  C>
  425  430  435  440  445  450  455  460  465  470  475  480  485  490
GGCGGTGCAGATTGTATGCCAACGGTAACACAACAAACGCCAACACCACCGATTTGGACGTTACCACCCA
 G  G  A  D  C  M  P  T  V  T  Q  Q  T  P  T  P  P  I  W  T  L  P  P>
  495  500  505  510  515  520  525  530  535  540  545  550  555  560
CGCAGACACCAGGATGGACTCCTGGACCACCGCTAACTCCAAAACCAACGACACCTCCGCATGTGACACC
T  Q  T  P  G  W  T  P  G  P  P  L  T  P  K  P  T  T  P  P  H  V  T  P>
  565  570  575  580  585  590  595  600  605  610  615  620  625  630
AGGAGGTTGTCGTTTGTGTCCATGTTACATACCTCCTCCATGTCAGATATGTCAACCATGTCAATGAGAA
  G  G  C  R  L  C  P  C  Y  I  P  P  P  C  Q  I  C  Q  P  C  Q  *>
  635  640  645  650  655  660
CTATGAGCACTTCATTCAACCATAGCTAAAGC
```

The coding sequence of *Bm-shp-1* including the SL1 leader sequence (in bold) was cloned previously by other researchers. Primers in turquoise were used to amplify a a

fragment used in the targeting construct. Primers in green were used to assess transcript levels in single worm RT-PCRs.

Bm-tub-1

```
   5   10   15   20   25   30   35   40   45   50   55   60   65   70
ATATGTGCCACGAGCAGTCCTTGTTGATTTGGAACCAGGTACCATGGATTCTATTCGAGGAGGTGAGTT

  75   80   85   90   95  100  105  110  115  120  125  130  135  140
GGGCAACTATTCCGACCTGACAATTTTGTTTTTGGGCAAAGTGGAGCTGGCAACAACTGGGCTAAGGGA

 145  150  155  160  165  170  175  180  185  190  195  200  205  210
ATTATACGGAAGGTGCGGAACTAGTTGATAATGTGTTGGACGTGATACGAAAAGAAGCTGAGGGATGCG

 215  220  225  230  235  240  245  250  255  260  265  270  275  280
TTGTCTTCAGGTACGGATTGCCATAGTTTTATAAGACATTTTAATTGTGGATGTGTTTTTATTTTAAAG

 285  290  295  300  305  310  315  320  325  330  335  340  345  350
AAAAATTAAAAAATATTTTCAGGGATTTCAACTAACGCATTCACTTGGTGGTGGTACCGGTTCCGGCAT

 355  360  365  370  375  380  385  390
GGAACATTGCTGATCTCGAAAATTCGTGAGGAGTATCCG
```

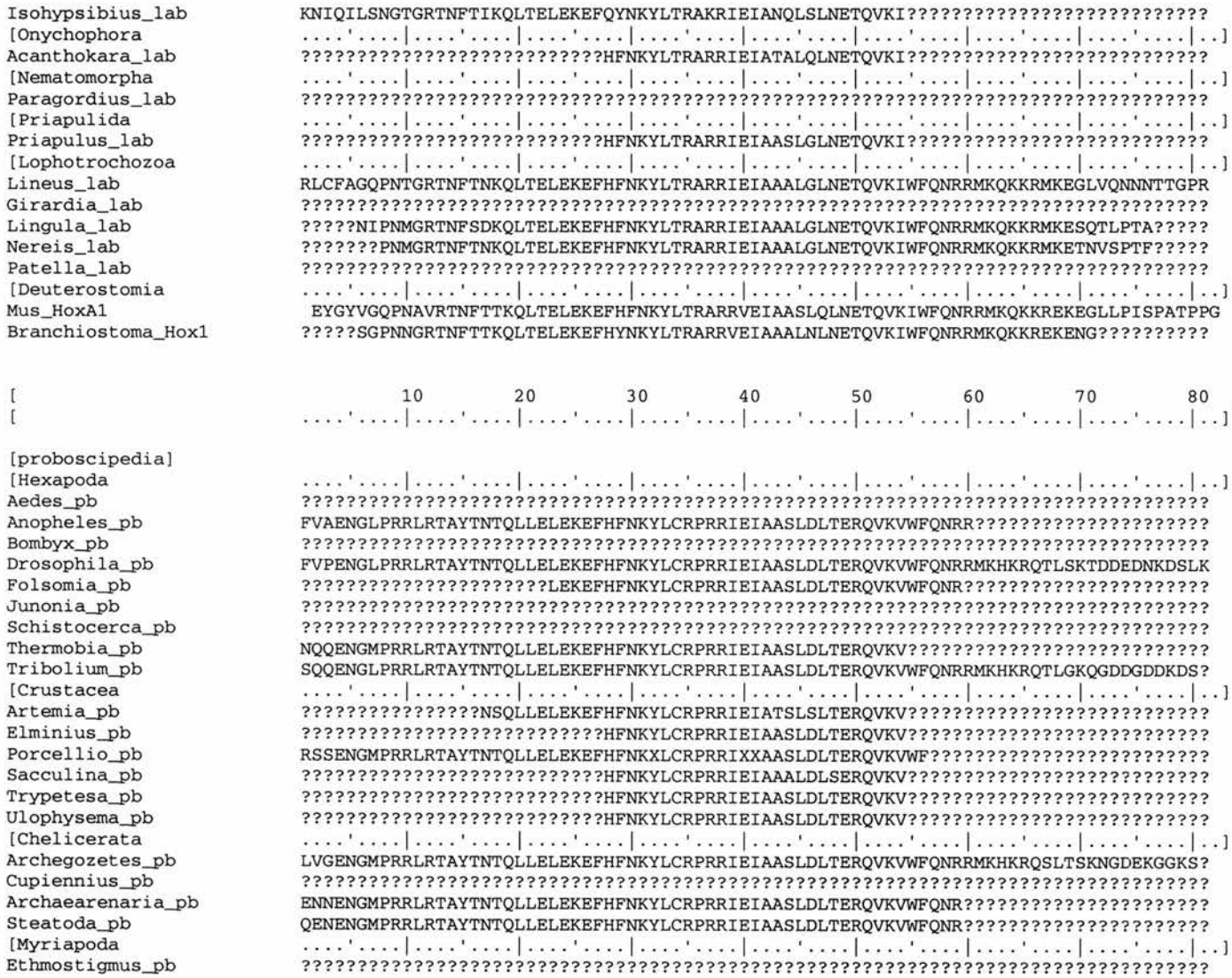
A fragment of the *B. malayi* β-tubulin gene including a small intron in yellow was amplified with a primer set that amplified a 390 bp fragment from genomic DNA and a 309 bp fragment from cDNA (primers in turquoise). The same primers were used to construct the targeting construct for generating dsRNA, assessing transcript levels from in single worms and to normalise cDNA levels across different lifecycle stages.

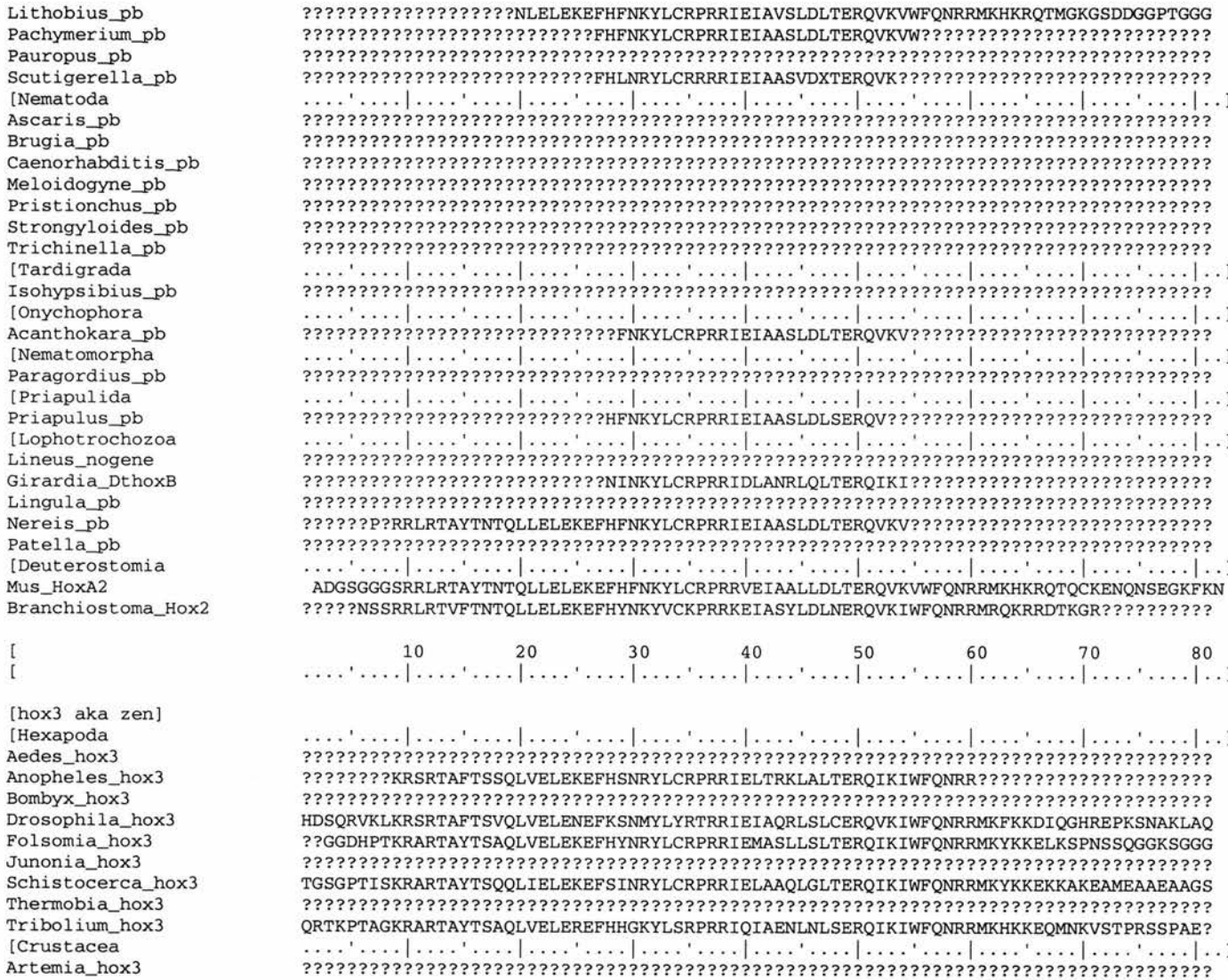
APPENDIX 3 ALLIGNMENT OF HOX GENES IN NEXUS FORMAT AND TABLE OF FULL SPECIES NAMES

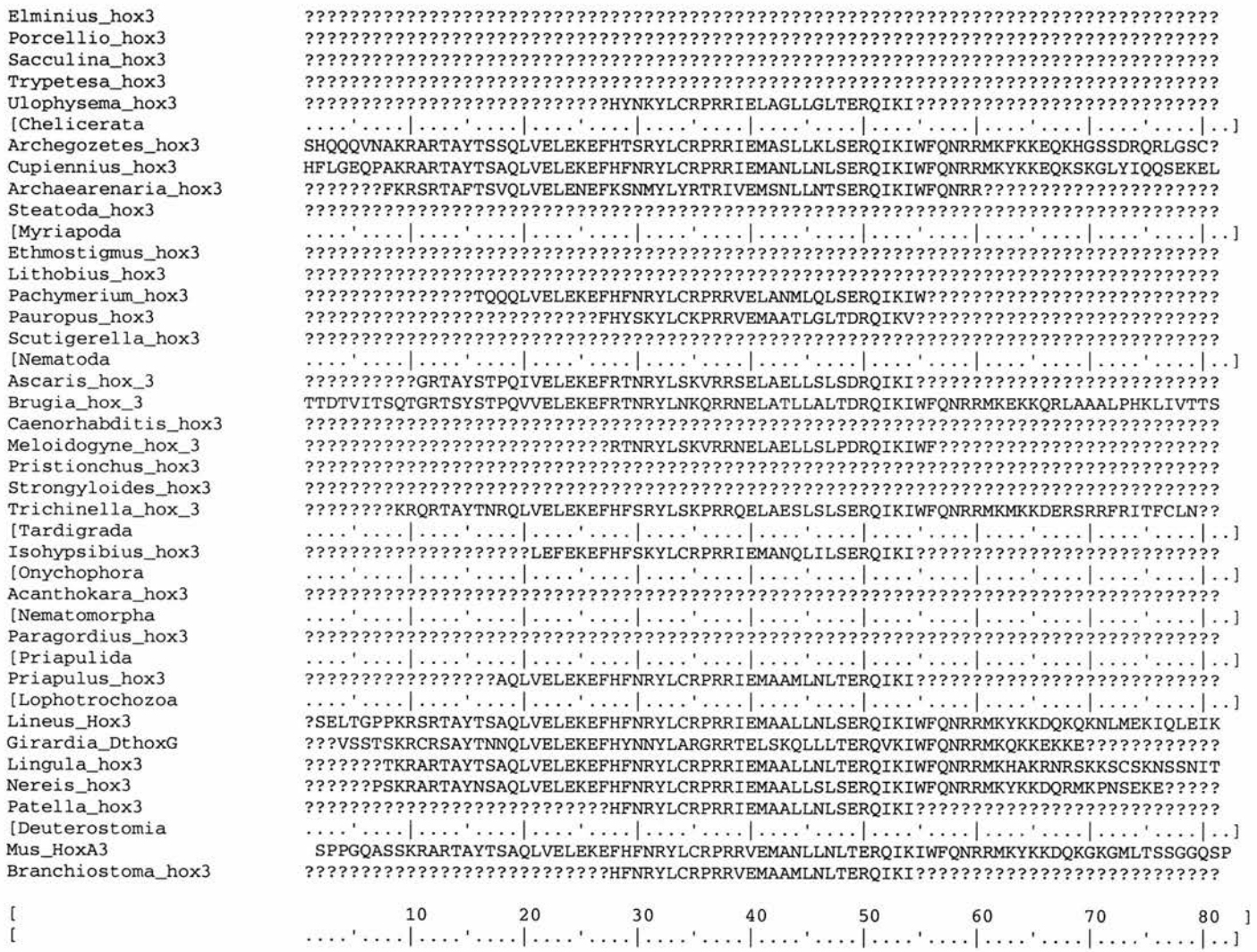
Allignment of selected Hox genes from the Bilateria

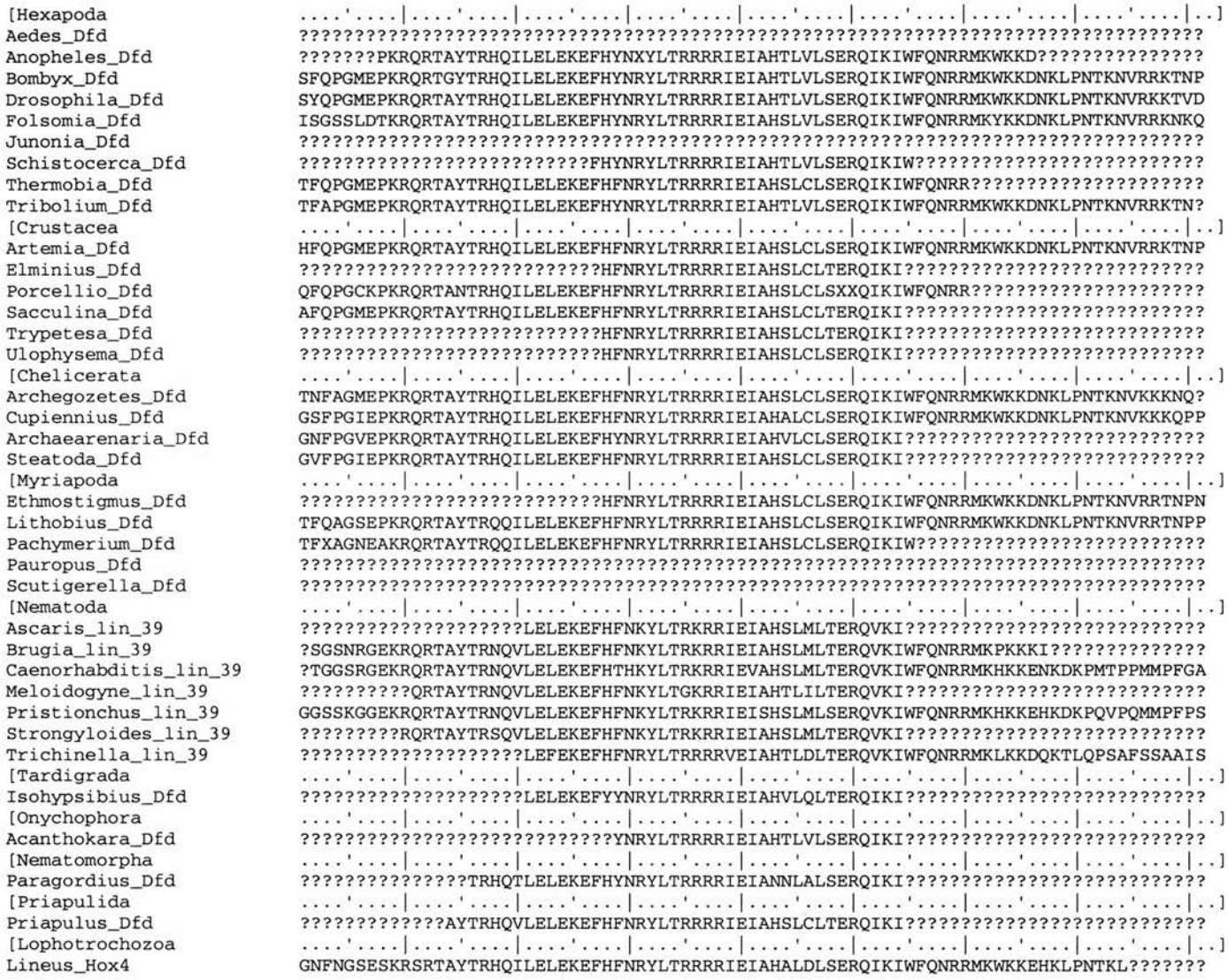
```
#NEXUS
[ecdysozoan and other HOX genes - arranged by species]
[Mark Blaxter and Aziz Aboobaker version 2  05.10.01]

[to convert to genes,
save the file as something new (eg HOX_meta_genes.nex"
find and replace "" with nothing, and "     " with nothing
find and replace "[" with "[", and ] with "]"
(this restores all the sequences to having unique gene names,
and brackets out the header file (replacing the "genes rather than species" one),
the gene members ship block and the assumptions block (replacing the "genes rather than species" one)]

[PLEASE NOTE!!!! You CANNOT GO BACK to this by-species arrangement from the by-genes format
so perform these operations on a copy of this file!!!!]

[alternate header for datafile as genes rather than species]

BEGIN DATA [block-name] ;
     DIMENSIONS NTAX=550 NCHAR=81;
     FORMAT
          MISSING=?
          SYMBOLS="0 1"
          INTERLEAVE
          DATATYPE = protein
          GAP=-;
     MATRIX


[alternate header for datafile as species]

[BEGIN DATA [block-name] ;
     DIMENSIONS NTAX=42 NCHAR=987;
     FORMAT
          MISSING=?
          SYMBOLS="0 1"
          INTERLEAVE
```

```
        DATATYPE = protein
        GAP=-;
    MATRIX  ]

[                              10        20        30        40        50        60        70        80  ]
[                        ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]

[labial characters 16-96]
[Hexapoda                ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_lab                ????????????????????????????????????????????????????????????????????????????????
Anopheles_lab            SSFMINNNSTGRTNFTNKQLTELEKEFHFNKYLTRARRIEIANALHLNETQVKIWFQNRRMKQK?????????????????
Bombyx_lab               ????????????????????????????????????????????????????????????????????????????????
Drosophila_lab           CSLSSNTNNSGRTNFTNKQLTELEKEFHFNRYLTRARRIEIANTLQLNETQVKIWFQNRRMKQKKRVKEGLIPADILTQHS
Folsomia_lab             ?GGGVGPGATGRTNFTTKQLTELEKEFHFNKYLTRARRIEIASSLQLNETQ????????????????????????????????
Junonia_lab              ????????????????????????????????????????????????????????????????????????????????
Schistocerca_lab         ??????????????????????????FHFNKYLTRARRIEIASALQLNETQVKIW?????????????????????????????
Thermobia_lab            VNGVGAGSNAGRTNFTNKQLTELEKEFHFNKYLTRARRIEIASALQLNETQVKIWFQNRR??????????????????????
Tribolium_lab            SMLSLNCLNTGRTNFTNKQLTELEKEFHFNKYLTRARRIEIASALQLNETQVKIWFQNRRMKQKKRMKEGLIPPEPISAS?
[Crustacea               ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_lab              ????????????????????????????????????????????????????????????????????????????????
Elminius_lab             ??????????????????????????HFNKYLTRARRIEIATALTLNETQVKI?????????????????????????????
Porcellio_lab            EYGGGFGGGTGRTNFTTKQLTELEKEFHFNKYLTRARRIEIASALQLNETQVKIWFQNRR??????????????????????
Sacculina_lab            ??????????????????????????HFNKYLTRARRIEIATALALNETQVKI?????????????????????????????
Trypetesa_lab            ??????????????????????????HFNKYLTRARRIEIAGALMLNETQVKI?????????????????????????????
Ulophysema_lab           ??????????????????????????HFNKYLTRARRIEIASALQLNETQVKI?????????????????????????????
[Chelicerata             ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_lab         ????????????????????????????????????????????????????????????????????????????????
Cupiennius_lab           NVSNGGSNGSGRTNFTTKQLTELEKEFHYNKYLTRARRIEIATALQLNETQVKIWFQNRRMKQKKRMKEGLLVTDVGSAKE
Archaearenaria_lab       ????????????????????????????????????????????????????????????????????????????????
Steatoda_lab             ??????????????????????????HFNKYLTRARRIEIASSLGLNETQVKI?????????????????????????????
[Myriapoda               ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_lab         ??????????????????????????HFNKYLTRARRIEIASALQLNETQVKI?????????????????????????????
Lithobius_lab            ??????????????????????????FHFNKYLTRARRIEIATALQLNETQVKI?????????????????????????????
Pachymerium_lab          ????????????????????????????????????????????????????????????????????????????????
Pauropus_lab             ??????????????????????????FHFNKYLTRVRRIEIASALQLNETQVKI?????????????????????????????
Scutigerella_lab         ??????????????????????????FHFNKYLTRARRIEIASTLQLNETQVKI?????????????????????????????
[Nematoda                ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_ceh_13           ???????????RTNFTTHQLTELEKEYYTSKYLNRTRRAEIASILQLNETQVKI?????????????????????????????
Brugia_ceh_13            KRRVVEDTNAIRTNFTTHQLTELEKEYYTSKYLNRTRRAEIASILQLNETQVKIWFQNRR??????????????????????
Caenorhabditis_ceh_13    KKKVIDENGTNRTNFTTHQLTELEKEFHTAKYVNRTRRTEIASNLKLQEAQVKIWFQNRRMKEKKREKEKAFLARNTWESN
Meloidogyne_ceh_13       ??????????????????????LELEKEFHTNKYLNKSRRAEIAAMLQLHESQIKIWF?????????????????????????
Pristionchus_ceh_13      ??????????????????????ELEKEFHTNKYVNRQRRTDIAAQLKLNEAQVKI?????????????????????????????
Strongyloides_ceh_13     ???????????RTNFTTHQLTELEKEYYTSKYLDRSRRREIAKQLALNETQVKI?????????????????????????????
Trichinella_ceh_13          LLNSGCTSVPNRTNFTTKQLTELEKEFHTNRYLTRARRIEIASQLGLNETQVKI???N?????????????????????????
[Tardigrada              ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
```

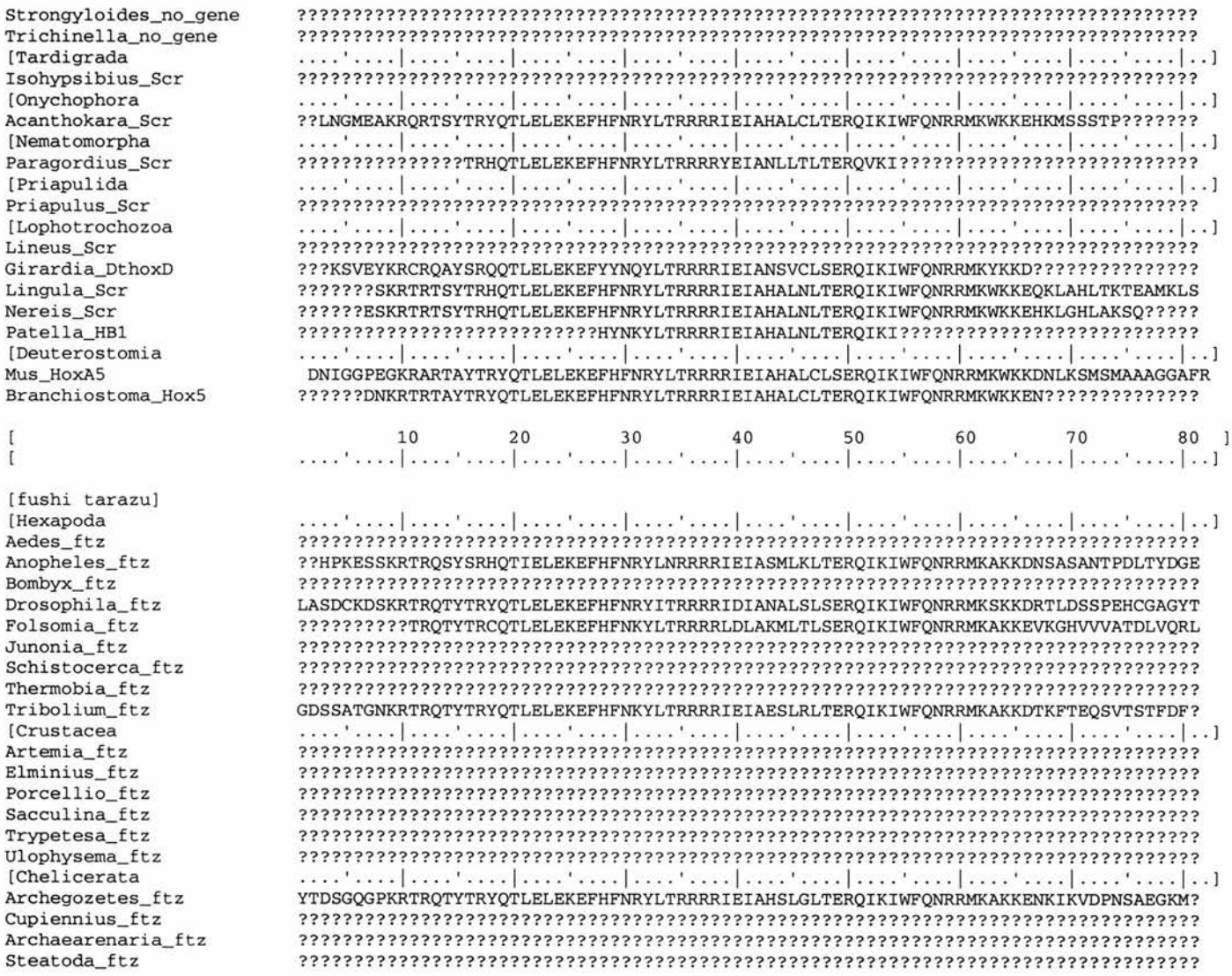

Isohypsibius_lab KNIQILSNGTGRTNFTIKQLTELEKEFQYNKYLTRAKRIEIANQLSLNETQVKI??????????????????????????
[Onychophora ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Acanthokara_lab ??????????????????????????HFNKYLTRARRIEIATALQLNETQVKI????????????????????????????
[Nematomorpha ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Paragordius_lab ??
[Priapulida ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Priapulus_lab ??????????????????????????????HFNKYLTRARRIEIAASLGLNETQVKI??????????????????????????
[Lophotrochozoa ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Lineus_lab RLCFAGQPNTGRTNFTNKQLTELEKEFHFNKYLTRARRIEIAAALGLNETQVKIWFQNRRMKQKKRMKEGLVQNNNTTGPR
Girardia_lab ??
Lingula_lab ?????NIPNMGRTNFSDKQLTELEKEFHFNKYLTRARRIEIAAALGLNETQVKIWFQNRRMKQKKRMKESQTLPTA?????
Nereis_lab ???????PNMGRTNFTNKQLTELEKEFHFNKYLTRARRIEIAAALGLNETQVKIWFQNRRMKQKKRMKETNVSPTF?????
Patella_lab ??
[Deuterostomia ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Mus_HoxA1 EYGYVGQPNAVRTNFTTKQLTELEKEFHFNKYLTRARRVEIAASLQLNETQVKIWFQNRRMKQKKREKEGLLPISPATPPG
Branchiostoma_Hox1 ?????SGPNNGRTNFTTKQLTELEKEFHYNKYLTRARRVEIAAALNLNETQVKIWFQNRRMKQKKREKENG??????????
[10 20 30 40 50 60 70 80]
[....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
[proboscipedia]
[Hexapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_pb ??
Anopheles_pb FVAENGLPRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNRR??????????????????????
Bombyx_pb ??
Drosophila_pb FVPENGLPRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNRRMKHKRQTLSKTDDEDNKDSLK
Folsomia_pb ????????????????????????LEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNR??????????????????????
Junonia_pb ??
Schistocerca_pb ??
Thermobia_pb NQQENGMPRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKV??????????????????????????
Tribolium_pb SQQENGLPRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNRRMKHKRQTLGKQGDDGDDKDS?
[Crustacea ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_pb ??????????????????NSQLLELEKEFHFNKYLCRPRRIEIATSLSLTERQVKV??????????????????????????????
Elminius_pb ??????????????????????????????HFNKYLCRPRRIEIAASLDLTERQVKV??????????????????????????
Porcellio_pb RSSENGMPRRLRTAYTNTQLLELEKEFHFNKXLCRPRRIXXAASLDLTERQVKVWF?????????????????????????
Sacculina_pb ????????????????????????????????HFNKYLCRPRRIEIAAALDLSERQVKV??????????????????????????
Trypetesa_pb ????????????????????????????????HFNKYLCRPRRIEIAASLDLTERQVKV??????????????????????????
Ulophysema_pb ????????????????????????????????HFNKYLCRPRRIEIAASLDLTERQVKV??????????????????????????
[Chelicerata ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_pb LVGENGMPRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNRRMKHKRQSLTSKNGDEKGGKS?
Cupiennius_pb ??
Archaearenaria_pb ENNENGMPRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNR??????????????????????
Steatoda_pb QENENGMPRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNR??????????????????????
[Myriapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_pb ??

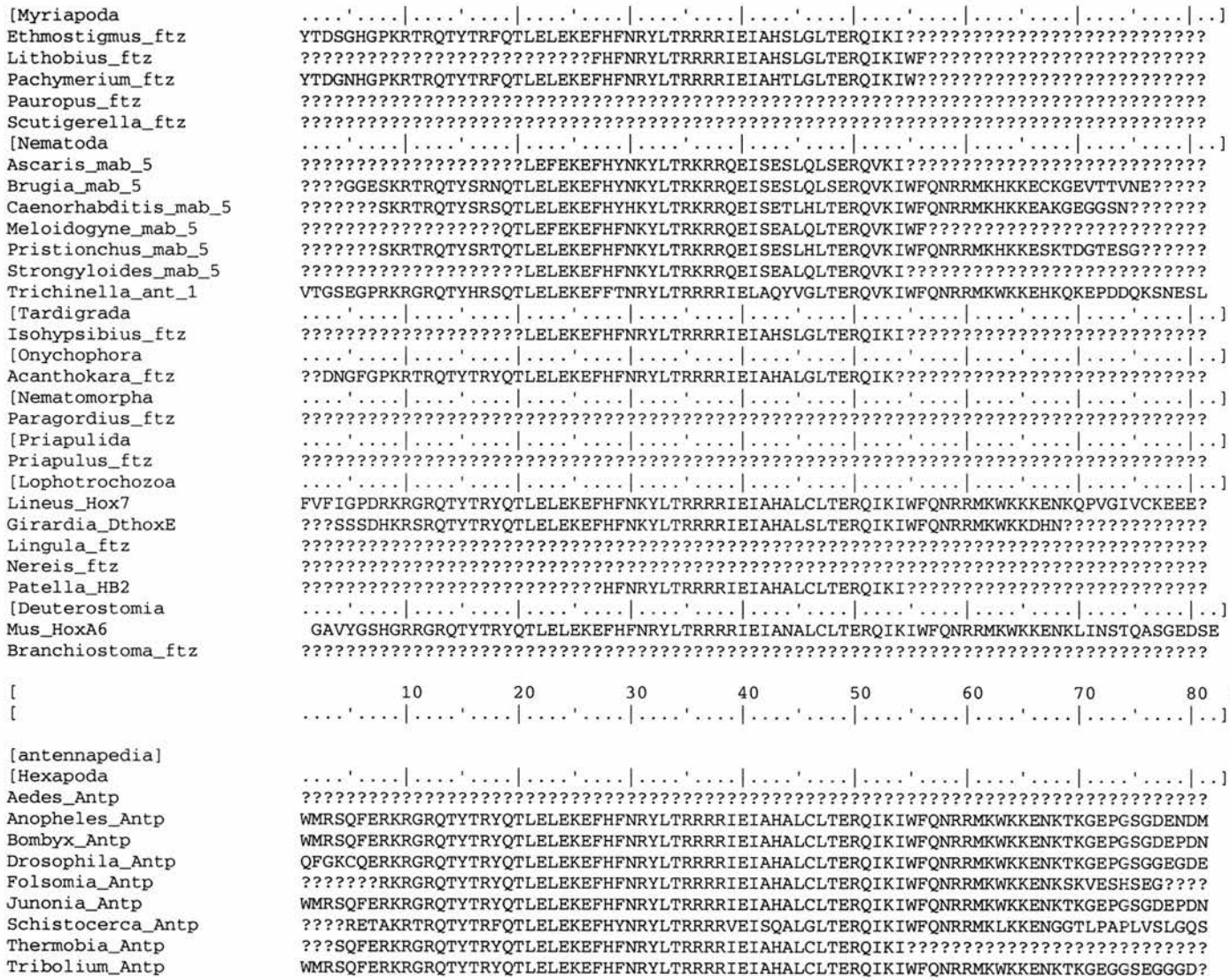

Lithobius_pb ??????????????????NLELEKEFHFNKYLCRPRRIEIAVSLDLTERQVKVWFQNRRMKHKRQTMGKGSDDGGPTGGG
Pachymerium_pb ??????????????????????????FHFNKYLCRPRRIEIAASLDLTERQVKVW????????????????????????
Pauropus_pb ??
Scutigerella_pb ??????????????????????????FHLNRYLCRRRRIEIAASVDXTERQVK????????????????????????????
[Nematoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_pb ??
Brugia_pb ??
Caenorhabditis_pb ??
Meloidogyne_pb ??
Pristionchus_pb ??
Strongyloides_pb ??
Trichinella_pb ??
[Tardigrada ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Isohypsibius_pb ??
[Onychophora ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Acanthokara_pb ??????????????????????????FNKYLCRPRRIEIAASLDLTERQVKV????????????????????????
[Nematomorpha ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Paragordius_pb ??
[Priapulida ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Priapulus_pb ??????????????????????????HFNKYLCRPRRIEIAASLDLSERQV????????????????????????????
[Lophotrochozoa ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Lineus_nogene ??
Girardia_DthoxB ??????????????????????????NINKYLCRPRRIDLANRLQLTERQIKI????????????????????????
Lingula_pb ??
Nereis_pb ??????P?RRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKV????????????????????????
Patella_pb ??
[Deuterostomia ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Mus_HoxA2 ADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKN
Branchiostoma_Hox2 ?????NSSRRLRTVFTNTQLLELEKEFHYNKYVCKPRRKEIASYLDLNERQVKIWFQNRRMRQKRRDTKGR??????????
[10 20 30 40 50 60 70 80]
[....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
[hox3 aka zen]
[Hexapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_hox3 ??
Anopheles_hox3 ????????KRSRTAFTSSQLVELEKEFHSNRYLCRPRRIELTRKLALTERQIKIWFQNRR????????????????????
Bombyx_hox3 ??
Drosophila_hox3 HDSQRVKLKRSRTAFTSVQLVELENEFKSNMYLYRTRRIEIAQRLSLCERQVKIWFQNRRMKFKKDIQGHREPKSNAKLAQ
Folsomia_hox3 ??GGDHPTKRARTAYTSAQLVELEKEFHYNRYLCRPRRIEMASLLSLTERQIKIWFQNRRMKYKKELKSPNSSQGGKSGGG
Junonia_hox3 ??
Schistocerca_hox3 TGSGPTISKRARTAYTSQQLIELEKEFSINRYLCRPRRIELAAQLGLTERQIKIWFQNRRMKYKKEKKAKEAMEAAEAAGS
Thermobia_hox3 ??
Tribolium_hox3 QRTKPTAGKRARTAYTSAQLVELEREFHHGKYLSRPRRIQIAENLNLSERQIKIWFQNRRMKHKKEQMNKVSTPRSSPAE?
[Crustacea ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_hox3 ??

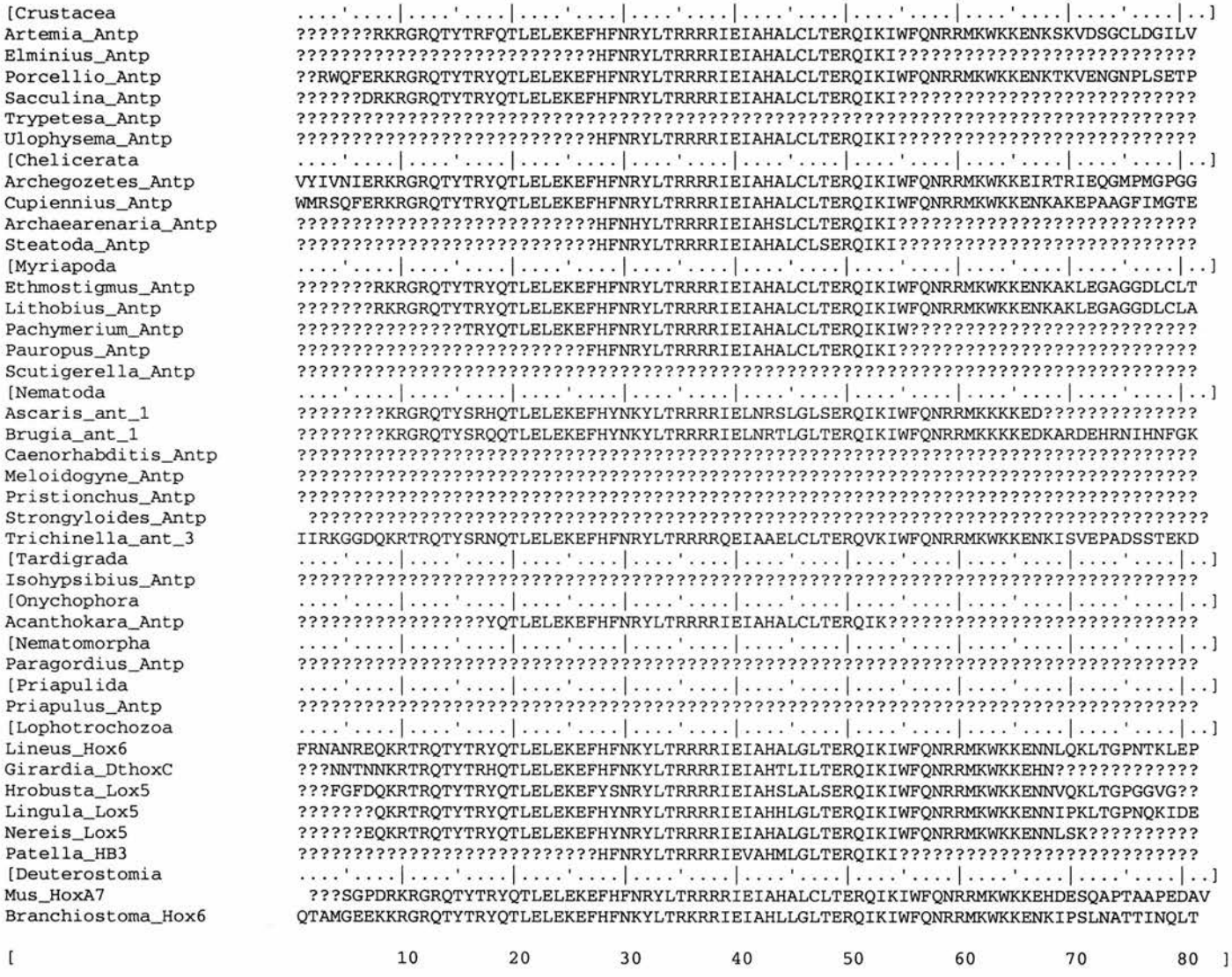

Elminius_hox3
Porcellio_hox3
Sacculina_hox3
Trypetesa_hox3
Ulophysema_hox3
[Chelicerata
Archegozetes_hox3
Cupiennius_hox3
Archaearenaria_hox3
Steatoda_hox3
[Myriapoda
Ethmostigmus_hox3
Lithobius_hox3
Pachymerium_hox3
Pauropus_hox3
Scutigerella_hox3
[Nematoda
Ascaris_hox_3
Brugia_hox_3
Caenorhabditis_hox3
Meloidogyne_hox_3
Pristionchus_hox3
Strongyloides_hox3
Trichinella_hox_3
[Tardigrada
Isohypsibius_hox3
[Onychophora
Acanthokara_hox3
[Nematomorpha
Paragordius_hox3
[Priapulida
Priapulus_hox3
[Lophotrochozoa
Lineus_Hox3
Girardia_DthoxG
Lingula_hox3
Nereis_hox3
Patella_hox3
[Deuterostomia
Mus_HoxA3
Branchiostoma_hox3
[10 20 30 40 50 60 70 80]
[deformed]

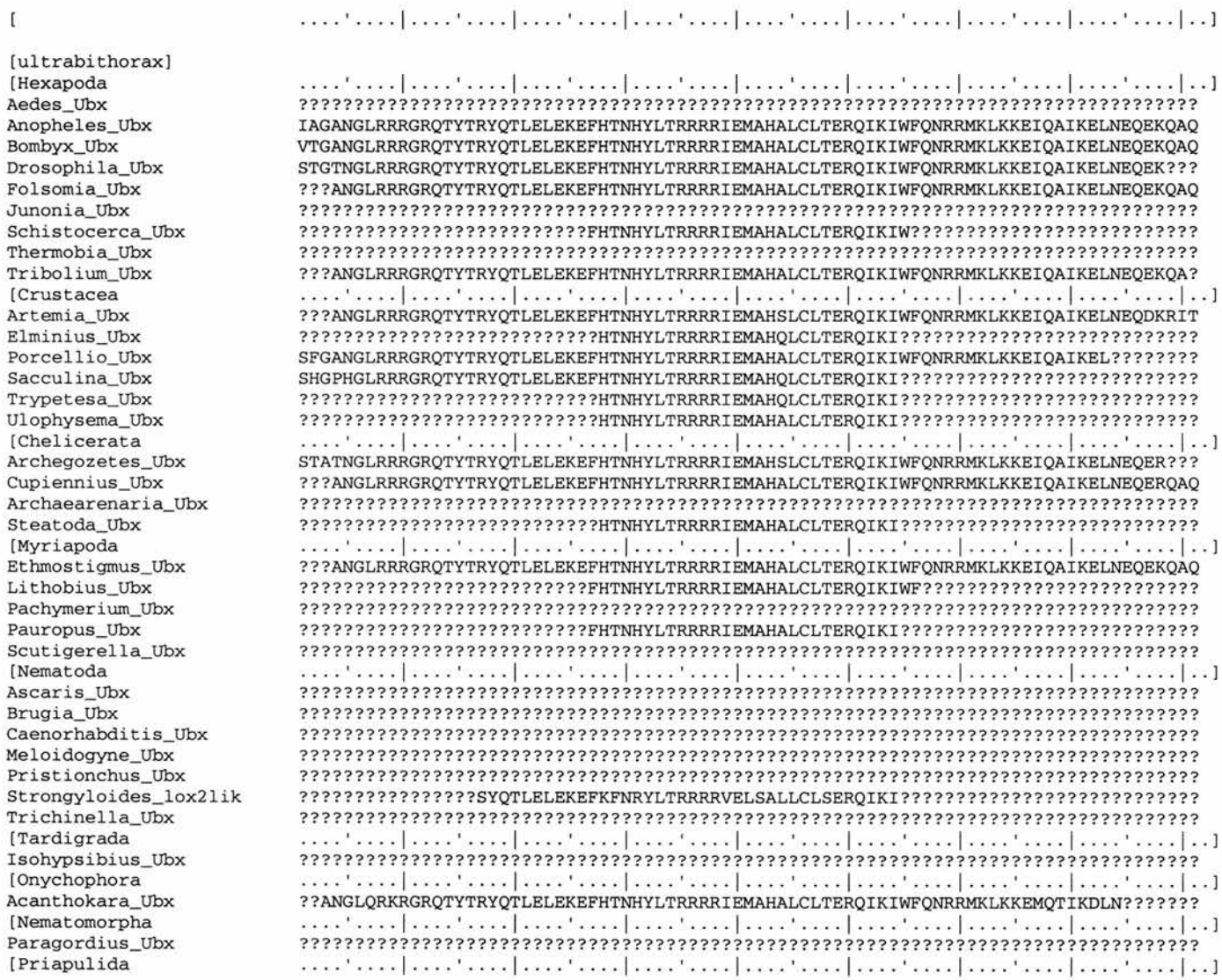

[Hexapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_Dfd ??
Anopheles_Dfd ???????PKRQRTAYTRHQILELEKEFHYNXYLTRRRRIEIAHTLVLSERQIKIWFQNRRMKWKKD??????????????
Bombyx_Dfd SFQPGMEPKRQRTGYTRHQILELEKEFHYNRYLTRRRRIEIAHTLVLSERQIKIWFQNRRMKWKKDNKLPNTKNVRRKTNP
Drosophila_Dfd SYQPGMEPKRQRTAYTRHQILELEKEFHYNRYLTRRRRIEIAHTLVLSERQIKIWFQNRRMKWKKDNKLPNTKNVRKKTVD
Folsomia_Dfd ISGSSLDTKRQRTAYTRHQILELEKEFHYNRYLTRRRRIEIAHSLVLSERQIKIWFQNRRMKYKKDNKLPNTKNVRRKNKQ
Junonia_Dfd ??
Schistocerca_Dfd ??????????????????????????FHYNRYLTRRRRIEIAHTLVLSERQIKIW????????????????????????????
Thermobia_Dfd TFQPGMEPKRQRTAYTRHQILELEKEFHFNRYLTRRRRIEIAHSLCLSERQIKIWFQNRR??????????????????????
Tribolium_Dfd TFAPGMEPKRQRTAYTRHQILELEKEFHYNRYLTRRRRIEIAHTLVLSERQIKIWFQNRRMKWKKDNKLPNTKNVRRKTN?
[Crustacea ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_Dfd HFQPGMEPKRQRTAYTRHQILELEKEFHFNRYLTRRRRIEIAHSLCLSERQIKIWFQNRRMKWKKDNKLPNTKNVRRKTNP
Elminius_Dfd ??????????????????????????HFNRYLTRRRRIEIAHSLCLTERQIKI??????????????????????????????
Porcellio_Dfd QFQPGCKPKRQRTANTRHQILELEKEFHFNRYLTRRRRIEIAHSLCLSXXQIKIWFQNRR??????????????????????
Sacculina_Dfd AFQPGMEPKRQRTAYTRHQILELEKEFHFNRYLTRRRRIEIAHSLCLTERQIKI??????????????????????????????
Trypetesa_Dfd ??????????????????????????????HFNRYLTRRRRIEIAHSLCLTERQIKI??????????????????????????????
Ulophysema_Dfd ??????????????????????????HFNRYLTRRRRIEIAHSLCLSERQIKI??????????????????????????????
[Chelicerata ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_Dfd TNFAGMEPKRQRTAYTRHQILELEKEFHFNRYLTRRRRIEIAHSLCLSERQIKIWFQNRRMKWKKDNKLPNTKNVKKKNQ?
Cupiennius_Dfd GSFPGIEPKRQRTAYTRHQILELEKEFHFNRYLTRRRRIEIAHALCLSERQIKIWFQNRRMKWKKDNKLPNTKNVKKKQPP
Archaearenaria_Dfd GNFPGVEPKRQRTAYTRHQILELEKEFHYNRYLTRRRRIEIAHVLCLSERQIKI??????????????????????????
Steatoda_Dfd GVFPGIEPKRQRTAYTRHQILELEKEFHFNRYLTRRRRIEIAHSLCLSERQIKI??????????????????????????
[Myriapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_Dfd ??????????????????????????HFNRYLTRRRRIEIAHSLCLSERQIKIWFQNRRMKWKKDNKLPNTKNVRRTNPN
Lithobius_Dfd TFQAGSEPKRQRTAYTRQQILELEKEFHFNRYLTRRRRIEIAHSLCLSERQIKIWFQNRRMKWKKDNKLPNTKNVRRTNPP
Pachymerium_Dfd TFXAGNEAKRQRTAYTRQQILELEKEFHFNRYLTRRRRIEIAHSLCLSERQIKIW????????????????????????
Pauropus_Dfd ??
Scutigerella_Dfd ??
[Nematoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_lin_39 ??????????????????????LELEKEFHFNKYLTRKRRIEIAHSLMLTERQVKI??????????????????????????
Brugia_lin_39 ?SGSNRGEKRQRTAYTRNQVLELEKEFHFNKYLTRKRRIEIAHSLMLTERQVKIWFQNRRMKPKKKI??????????????
Caenorhabditis_lin_39 ?TGGSRGEKRQRTAYTRNQVLELEKEFHTHKYLTRKRRIEVAHSLMLTERQVKIWFQNRRMKHKKENKDKPMTPPMMPFGA
Meloidogyne_lin_39 ??????????QRTAYTRNQVLELEKEFHFNKYLTGKRRIEIAHTLILTERQVKI??????????????????????????
Pristionchus_lin_39 GGSSKGGEKRQRTAYTRNQVLELEKEFHFNKYLTRKRRIEISHSLMLSERQVKIWFQNRRMKHKKEHKDKPQVPQMMPFPS
Strongyloides_lin_39 ?????????RQRTAYTRSQVLELEKEFHFNKYLTRKRRIEIAHSLMLTERQVKI??????????????????????????
Trichinella_lin_39 ??????????????????????LEFEKEFHFNRYLTRRRVEIAHTLDLTERQVKIWFQNRRMKLKKDQKTLQPSAFSSAAIS
[Tardigrada ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Isohypsibius_Dfd ??????????????????????LELEKEFYYNRYLTRRRRIEIAHVLQLTERQIKI??????????????????????????
[Onychophora ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Acanthokara_Dfd ??????????????????????????????YNRYLTRRRRIEIAHTLVLSERQIKI??????????????????????????
[Nematomorpha ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Paragordius_Dfd ??????????????TRHQTLELEKEFHYNRYLTRRRRIEIANNLALSERQIKI??????????????????????????
[Priapulida ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Priapulus_Dfd ????????????AYTRHQVLELEKEFHFNRYLTRRRRIEIAHSLCLTERQIKI??????????????????????????
[Lophotrochozoa ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Lineus_Hox4 GNFNGSESKRSRTAYTRHQILELEKEFHFNRYLTRRRRIEIAHALDLSERQIKIWFQNRRMKWKKEHKLPNTKL???????

```
Girardia_DthoxA          ??????DSKRNRTAYTRQQILELEKEFHFNKYLTRKRRNEIAQSLQLSESQVKIWFQNRRMKWKKDHHL???????????
Lingula_Dfd              ????????????????????????????????????????????????????????????????????????????????
Nereis_Dfd               ??????DSKRTRTAYTRHQVLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKLPNTKNRL?????
Patella_Dfd              ????????????????????????????????????????????????????????????????????????????????
[Deuterostomia           ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Mus_HoxA4                 SSYNGGEPKRSRTAYTRQQVLELEKEFHFNRYLTRRRRIEIAHTLCLSERQVKIWFQNRRMKWKKDHLPNTKMRSSNTASA
Branchiostoma_Hox4       ?????QDTKRSRTAYTRQQVLELEKEFHFNRYLTRRRRIEIAHSLGLTERQIKIWFQNRRMKWKKDNRLPN??????????

[                                10        20        30        40        50        60        70        80  ]
[                        ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]

[sex combs reduced]
[Hexapoda                ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_Scr                ????????????????????????????????????????????????????????????????????????????????
Anopheles_Scr            TVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMASMNIVPYHMSP
Bombyx_Scr               TVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMASMNIVPYHMNP
Drosophila_Scr           TVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMASMNIVPYHMGP
Folsomia_Scr             TVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHSLCFSERQIKIWFQNRRMKWKKEHKMATMNALPHHHGM
Junonia_Scr              ????????????????????????EFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMASMNIVPYHMNP
Schistocerca_Scr         TVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMASMSAGGNYYAG
Thermobia_Scr            TVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Tribolium_Scr            TVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMASMNIVPYHMS?
[Crustacea               ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_Scr              AVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHTLCLSERQIKIWFQNRRMKWKKEHKIATMNMMQHHHHQ
Elminius_Scr             ????????????????????????????????????????????????????????????????????????????????
Porcellio_Scr            TVNSNGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Sacculina_Scr            QVNANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Trypetesa_Scr            ?VNSNGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Ulophysema_Scr           ????????????????????????????????????????????????????????????????????????????????
[Chelicerata             ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_Scr         DGVNAMGETKRQRTSYTRYQTLLEKEFHFNRYLTRRRRIEIAHSLCLSERQIKIWFQNRRMKWKKEHKMQSVVPPQIPQM?
Cupiennius_Scr           ????????????????????????????HFNQYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Archaearenaria_Scr       ????????????????????????????HFNQYLTRRRRIEIAHSLCLTERQIKI????????????????????????????
Steatoda_Scr             ????????????????????????????HFNQYLTRRRRIEIAHALCLTERQIKI????????????????????????????
[Myriapoda               ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_Scr         ????????????SYTRYQTLELEKEFHFNRYLTRRRRIEIAHSLCLSERQIKIWFQNRRMKWKKEHKIPSVNQVPGPIVH
Lithobius_Scr            ???????TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIW????????????????????????????
Pachymerium_Scr          ????????????????TRYQTLELEKEFHFNRYLTRRRRIEIAHALCLSERQIKIW????????????????????????????
Pauropus_Scr             ??????????????????????????FHFNRYLTRRRRIEIAHSLCLSERQIKI????????????????????????????
Scutigerella_Scr         ??????????????????????????FHFNRYLTRRRRIEIAHSLCLTERQIKI????????????????????????????
[Nematoda                ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_no_gene          ????????????????????????????????????????????????????????????????????????????????
Brugia_no_gene           ????????????????????????????????????????????????????????????????????????????????
Caenorhabditis_Dfd       ????????????????????????????????????????????????????????????????????????????????
Meloidogyne_no_gene      ????????????????????????????????????????????????????????????????????????????????
Pristionchus_no_gene     ????????????????????????????????????????????????????????????????????????????????
```

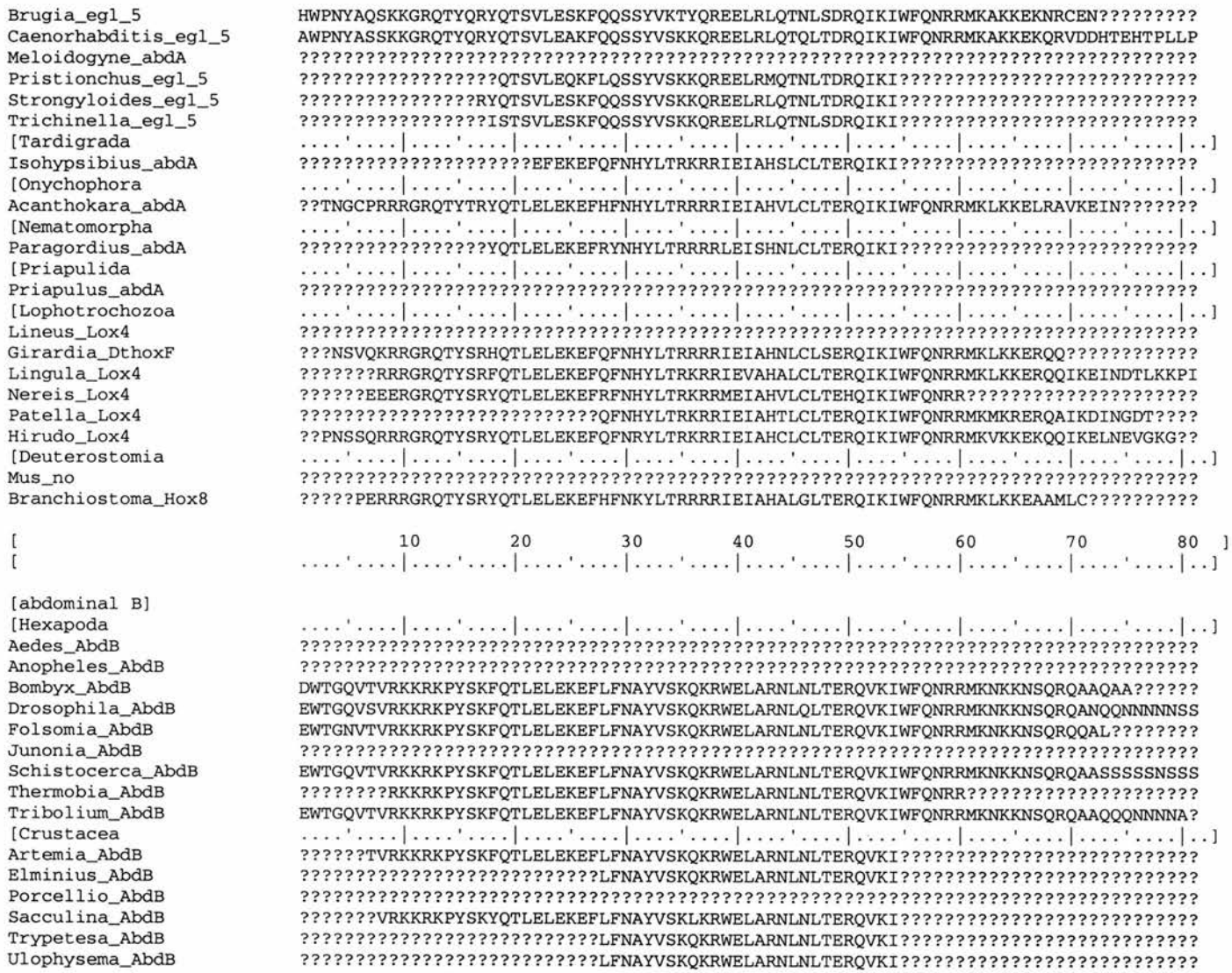

Strongyloides_no_gene
Trichinella_no_gene
[Tardigrada
Isohypsibius_Scr
[Onychophora
Acanthokara_Scr
??LNGMEAKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHKMSSSTP???????
[Nematomorpha
Paragordius_Scr
[Priapulida
Priapulus_Scr
[Lophotrochozoa
Lineus_Scr
Girardia_DthoxD
???KSVEYKRCRQAYSRQQTLELEKEFYYNQYLTRRRRIEIANSVCLSERQIKIWFQNRRMKYKKD???????????????
Lingula_Scr
???????SKRTRTSYTRHQTLELEKEFHFNRYLTRRRRIEIAHALNLTERQIKIWFQNRRMKWKKEQKLAHLTKTEAMKLS
Nereis_Scr
??????ESKRTRTSYTRHQTLELEKEFHFNRYLTRRRRIEIAHALNLTERQIKIWFQNRRMKWKKEHKLGHLAKSQ?????
Patella_HB1
????????????????????????????HYNKYLTRRRRIEIAHALNLTERQIKI???????????????????????????
[Deuterostomia
Mus_HoxA5
DNIGGPEGKRARTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLSERQIKIWFQNRRMKWKKDNLKSMSMAAAGGAFR
Branchiostoma_Hox5
??????DNKRTRTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN?????????????
[
10
20
30
40
50
60
70
80
]
[
[fushi tarazu]
[Hexapoda
Aedes_ftz
Anopheles_ftz
??HPKESSKRTRQSYSRHQTIELEKEFHFNRYLNRRRRIEIASMLKLTERQIKIWFQNRRMKAKKDNSASANTPDLTYDGE
Bombyx_ftz
Drosophila_ftz
LASDCKDSKRTRQTYTRYQTLELEKEFHFNRYITRRRRIDIANALSLSERQIKIWFQNRRMKSKKDRTLDSSPEHCGAGYT
Folsomia_ftz
??????????TRQTYTRCQTLELEKEFHFNKYLTRRRRLDLAKMLTLSERQIKIWFQNRRMKAKKEVKGHVVVATDLVQRL
Junonia_ftz
Schistocerca_ftz
Thermobia_ftz
Tribolium_ftz
GDSSATGNKRTRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAESLRLTERQIKIWFQNRRMKAKKDTKFTEQSVTSTFDF?
[Crustacea
Artemia_ftz
Elminius_ftz
Porcellio_ftz
Sacculina_ftz
Trypetesa_ftz
Ulophysema_ftz
[Chelicerata
Archegozetes_ftz
YTDSGQGPKRTRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHSLGLTERQIKIWFQNRRMKAKKENKIKVDPNSAEGKM?
Cupiennius_ftz
Archaearenaria_ftz
Steatoda_ftz

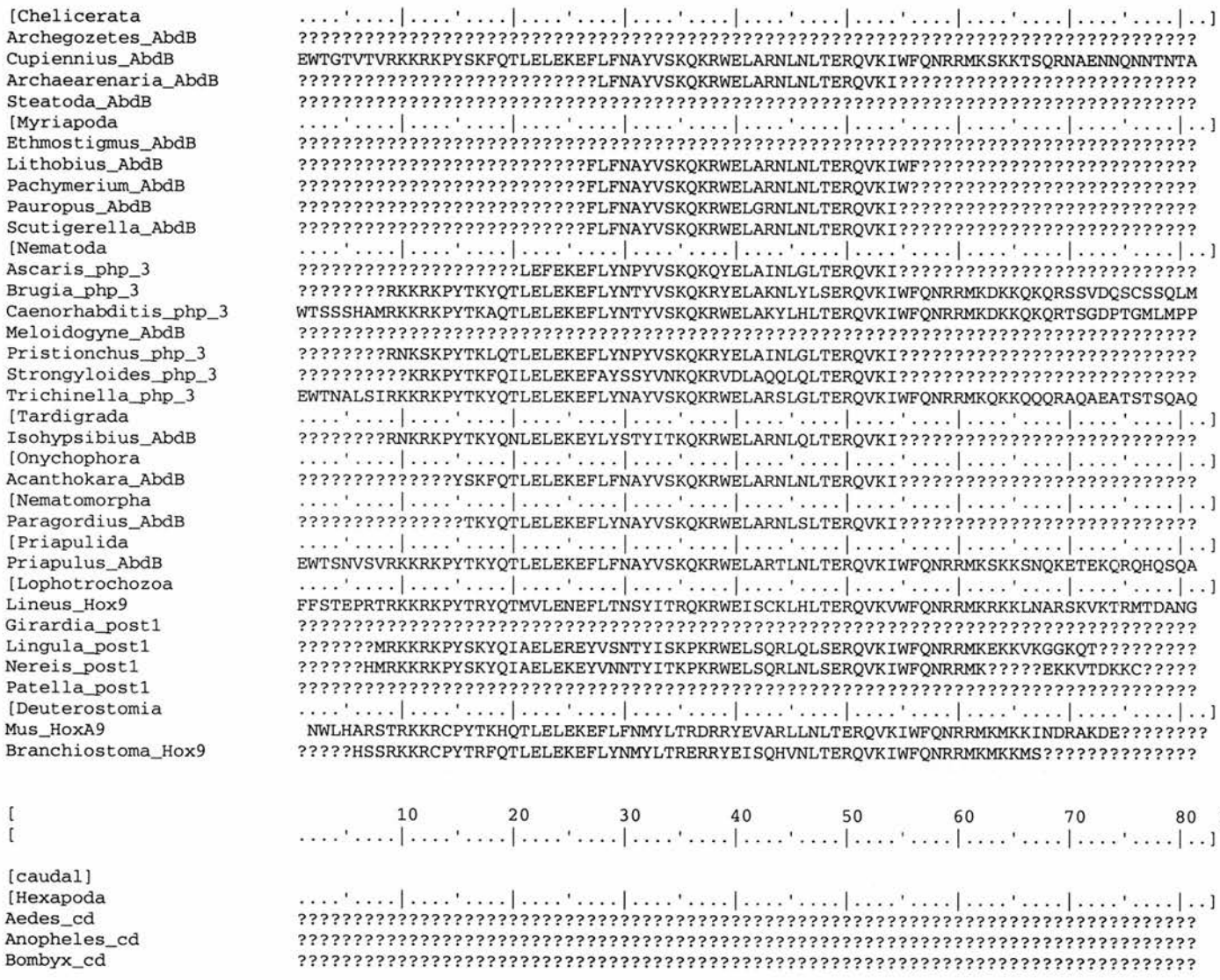

[Myriapoda
Ethmostigmus_ftz
Lithobius_ftz
Pachymerium_ftz
Pauropus_ftz
Scutigerella_ftz
[Nematoda
Ascaris_mab_5
Brugia_mab_5
Caenorhabditis_mab_5
Meloidogyne_mab_5
Pristionchus_mab_5
Strongyloides_mab_5
Trichinella_ant_1
[Tardigrada
Isohypsibius_ftz
[Onychophora
Acanthokara_ftz
[Nematomorpha
Paragordius_ftz
[Priapulida
Priapulus_ftz
[Lophotrochozoa
Lineus_Hox7
Girardia_DthoxE
Lingula_ftz
Nereis_ftz
Patella_HB2
[Deuterostomia
Mus_HoxA6
Branchiostoma_ftz
10
20
30
40
50
60
70
80
[antennapedia]
[Hexapoda
Aedes_Antp
Anopheles_Antp
Bombyx_Antp
Drosophila_Antp
Folsomia_Antp
Junonia_Antp
Schistocerca_Antp
Thermobia_Antp
Tribolium_Antp

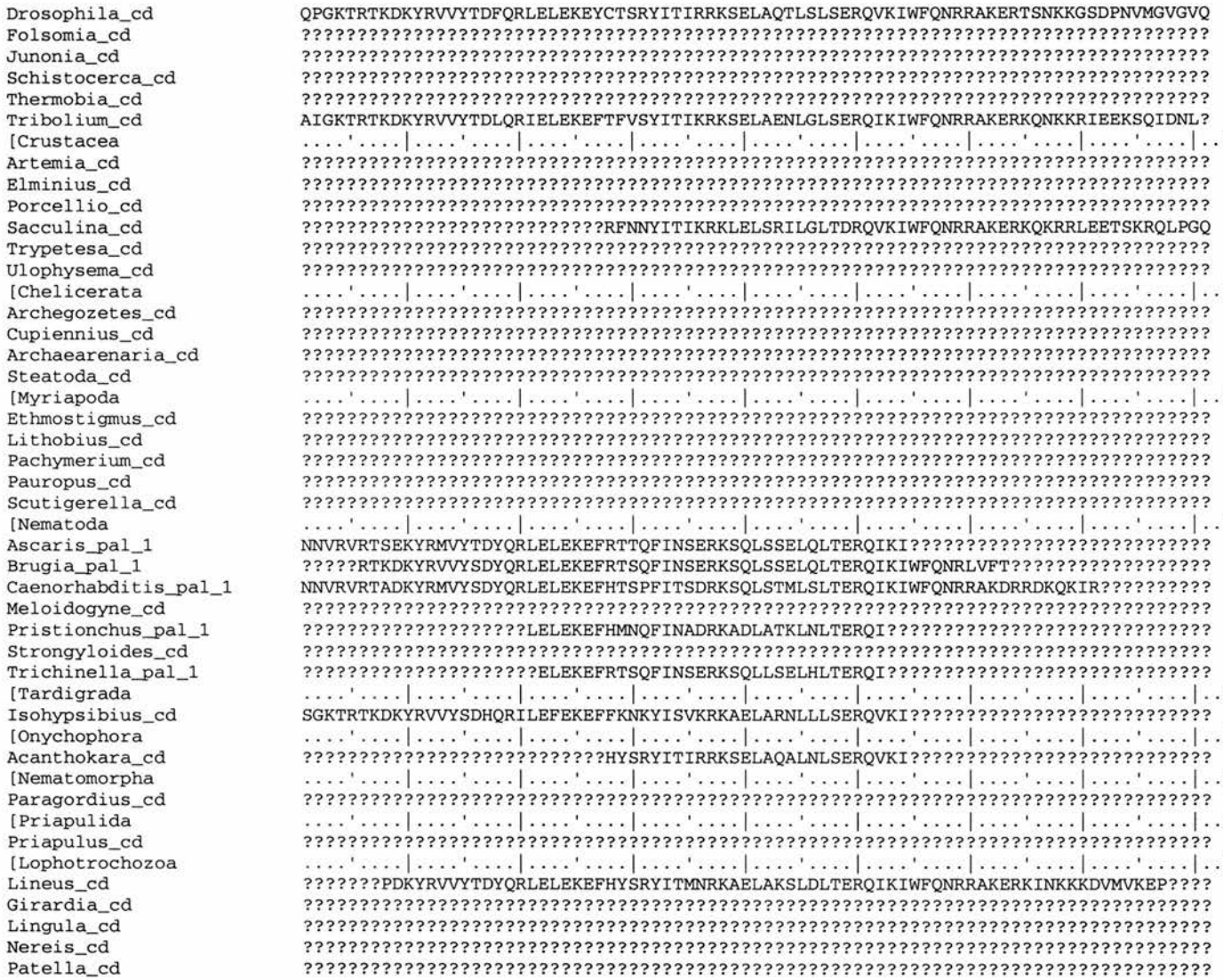

[Crustacea ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_Antp ???????RKRGRQTYTRFQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKSKVDSGCLDGILV
Elminius_Antp ??????????????????????????????HFNRYLTRRRRIEIAHALCLTERQIKI??????????????????????????????
Porcellio_Antp ??RWQFERKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKTKVENGNPLSETP
Sacculina_Antp ??????DRKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Trypetesa_Antp ??
Ulophysema_Antp ??????????????????????????????HFNRYLTRRRRIEIAHALCLTERQIKI????????????????????????????
[Chelicerata ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_Antp VYIVNIERKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEIRTRIEQGMPMGPGG
Cupiennius_Antp WMRSQFERKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKAKEPAAGFIMGTE
Archaearenaria_Antp ????????????????????????????HFNHYLTRRRRIEIAHSLCLTERQIKI????????????????????????????
Steatoda_Antp ????????????????????????????HFNRYLTRRRRIEIAHALCLSERQIKI????????????????????????????
[Myriapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_Antp ???????RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKAKLEGAGGDLCLT
Lithobius_Antp ???????RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKAKLEGAGGDLCLA
Pachymerium_Antp ????????????????TRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIW????????????????????????
Pauropus_Antp ??????????????????????????????FHFNRYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Scutigerella_Antp ??
[Nematoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_ant_1 ????????KRGRQTYSRHQTLELEKEFHYNKYLTRRRRIELNRSLGLSERQIKIWFQNRRMKKKKED?????????????
Brugia_ant_1 ????????KRGRQTYSRQQTLELEKEFHYNKYLTRRRRIELNRTLGLTERQIKIWFQNRRMKKKKEDKARDEHRNIHNFGK
Caenorhabditis_Antp ??
Meloidogyne_Antp ??
Pristionchus_Antp ??
Strongyloides_Antp ??
Trichinella_ant_3 IIRKGGDQKRTRQTYSRNQTLELEKEFHFNRYLTRRRQEIAAELCLTERQVKIWFQNRRMKWKKENKISVEPADSSTEKD
[Tardigrada ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Isohypsibius_Antp ??
[Onychophora ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Acanthokara_Antp ??????????????????YQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIK??????????????????????????
[Nematomorpha ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Paragordius_Antp ??
[Priapulida ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Priapulus_Antp ??
[Lophotrochozoa ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Lineus_Hox6 FRNANREQKRTRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAHALGLTERQIKIWFQNRRMKWKKENNLQKLTGPNTKLEP
Girardia_DthoxC ???NNTNNKRTRQTYTRHQTLELEKEFHFNKYLTRRRRIEIAHTLILTERQIKIWFQNRRMKWKKEHN?????????????
Hrobusta_Lox5 ???FGFDQKRTRQTYTRYQTLELEKEFYSNRYLTRRRRIEIAHSLALSERQIKIWFQNRRMKWKKENNVQKLTGPGGVG??
Lingula_Lox5 ???????QKRTRQTYTRYQTLELEKEFHYNRYLTRRRRIEIAHHLGLTERQIKIWFQNRRMKWKKENNIPKLTGPNQKIDE
Nereis_Lox5 ??????EQKRTRQTYTRYQTLELEKEFHYNRYLTRRRRIEIAHALGLTERQIKIWFQNRRMKWKKENNLSK??????????
Patella_HB3 ??????????????????????????????HFNRYLTRRRRIEVAHMLGLTERQIKI????????????????????????????
[Deuterostomia ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Mus_HoxA7 ???SGPDRKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEHDESQAPTAAPEDAV
Branchiostoma_Hox6 QTAMGEEKKRGRQTYTRYQTLELEKEFHFNKYLTRKRRIEIAHLLGLTERQIKIWFQNRRMKWKKENKIPSLNATTINQLT
[10 20 30 40 50 60 70 80]


[....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
[ultrabithorax]
[Hexapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_Ubx ??
Anopheles_Ubx IAGANGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQAQ
Bombyx_Ubx VTGANGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQAQ
Drosophila_Ubx STGTNGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEK???
Folsomia_Ubx ???ANGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQAQ
Junonia_Ubx ??
Schistocerca_Ubx ??????????????????????????FHTNHYLTRRRRIEMAHALCLTERQIKIW????????????????????????
Thermobia_Ubx ??
Tribolium_Ubx ???ANGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQA?
[Crustacea ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_Ubx ???ANGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHSLCLTERQIKIWFQNRRMKLKKEIQAIKELNEQDKRIT
Elminius_Ubx ??????????????????????????HTNHYLTRRRRIEMAHQLCLTERQIKI????????????????????????????
Porcellio_Ubx SFGANGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKEL????????
Sacculina_Ubx SHGPHGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHQLCLTERQIKI????????????????????????????
Trypetesa_Ubx ??????????????????????????HTNHYLTRRRRIEMAHQLCLTERQIKI????????????????????????????
Ulophysema_Ubx ??????????????????????????HTNHYLTRRRRIEMAHALCLTERQIKI????????????????????????????
[Chelicerata ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_Ubx STATNGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHSLCLTERQIKIWFQNRRMKLKKEIQAIKELNEQER???
Cupiennius_Ubx ???ANGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQERQAQ
Archaearenaria_Ubx ??
Steatoda_Ubx ??????????????????????????HTNHYLTRRRRIEMAHALCLTERQIKI????????????????????????????
[Myriapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_Ubx ???ANGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQAQ
Lithobius_Ubx ??????????????????????????FHTNHYLTRRRRIEMAHALCLTERQIKIWF????????????????????????????
Pachymerium_Ubx ??
Pauropus_Ubx ??????????????????????????FHTNHYLTRRRRIEMAHALCLTERQIKI????????????????????????????
Scutigerella_Ubx ??
[Nematoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_Ubx ??
Brugia_Ubx ??
Caenorhabditis_Ubx ??
Meloidogyne_Ubx ??
Pristionchus_Ubx ??
Strongyloides_lox2lik ???????????????????SYQTLELEKEFKFNRYLTRRRVELSALLCLSERQIKI????????????????????????
Trichinella_Ubx ??
[Tardigrada ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Isohypsibius_Ubx ??
[Onychophora ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Acanthokara_Ubx ??ANGLQRKRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEMQTIKDLN???????
[Nematomorpha ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Paragordius_Ubx ??
[Priapulida ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]

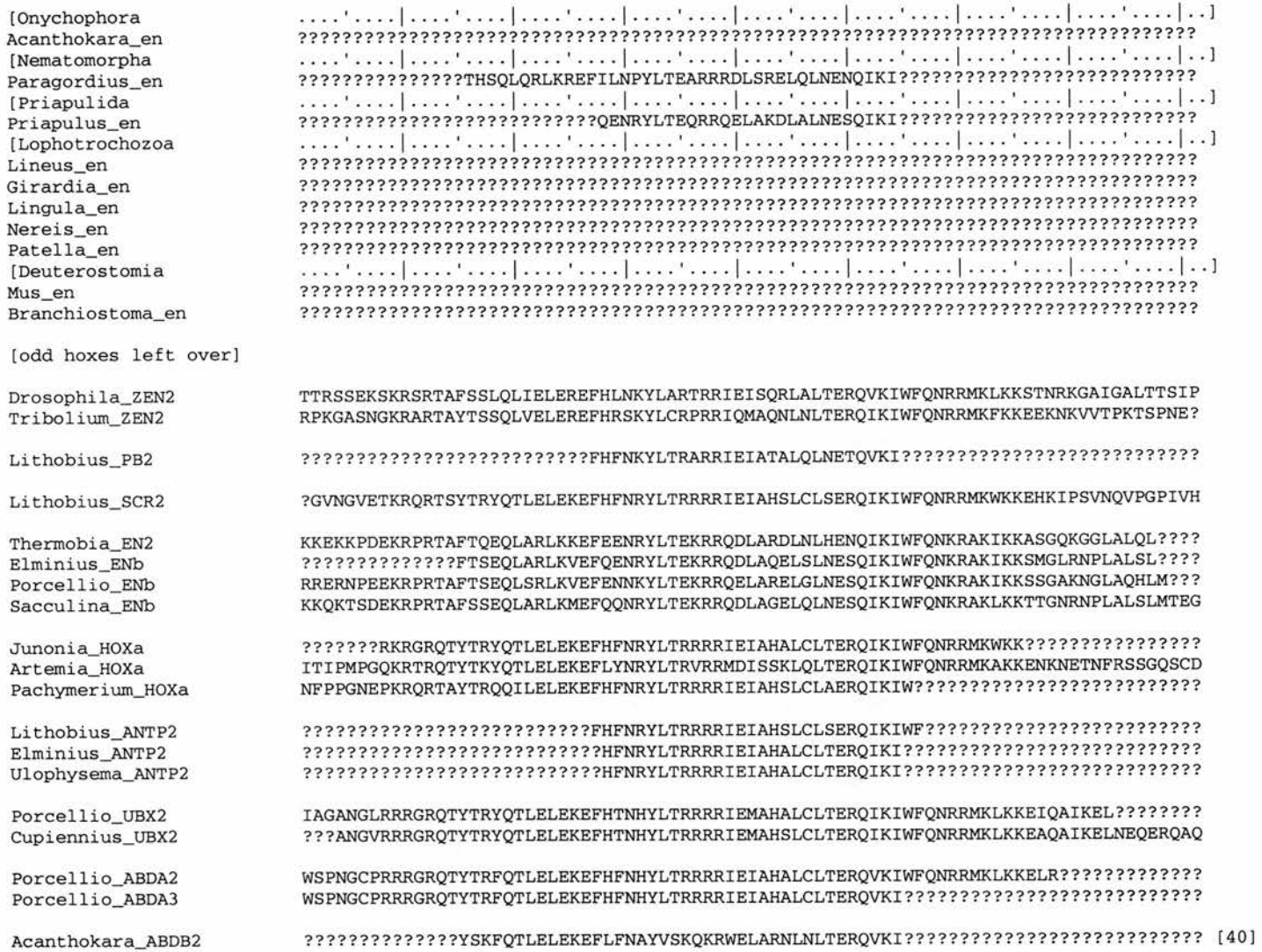
```
Priapulus_Ubx              ??ANGLFRRRGRQTYTRYQTLELEKEFRFNHYLTRRRRIEMSQALCLTERQIKIWFQNRRMKLKKETQALKEMNAHAKAL?
[Lophotrochozoa            ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Lineus_Lox2                ????????????????????????????????????????????????????????????????????????????????
Girardia_Ubx               ????????????????????????????????????????????????????????????????????????????????
Lingula_Lox2               ??????????????????????????????FNRYLTRRRRIELSHMLCLTERQI????????????????????????????
Nereis_Lox2                ??????????????????????????????KFNRYLTRKRRIELSHMLCLTERQIK????????????????????????????
Patella_Lox2               ??????????????????????????????KFNRYLTRRRRIELSHMLCLTERQIKIWFQNRRMKEKKELQAIKELNAQT????
Hirudo_Lox2                ??PNSNQRRRGRQTYTRYQTLELEKEFKFNRYLTRRRRIELSHTLYLTERQIKIWFQNRRMKEKKEVQAIRELNEIEKT??
[Deuterostomia             ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Mus_HoxB8                   RPQAAAGRRRGRQTYSRYQTLELEKEFLFNPYLTRKRRIEVSHALGLTERQVKIWFQNRRMKWKKENNKDKFPSSKCEQEE
Branchiostoma_Hox7         ?????PERKRGRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKLES??????????

[                                  10        20        30        40        50        60        70        80  ]
[                          ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]

[abdominal A]
[Hexapoda                  ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_abdA                 ??PNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQVSWRF
Anopheles_abdA             AGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQARRER
Bombyx_abdA                PGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQARRER
Drosophila_abdA            NGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQARRDR
Folsomia_abdA              ????GCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHAXCLTERQIKIWFQNRRMKLKKELRAVKEINEQARRER
Junonia_abdA               ????????????????????????????EFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQARRER
Schistocerca_abdA          ??PNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQARRER
Thermobia_abdA             CGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLPERQIKI????????????????????????????
Tribolium_abdA             CGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQARRE?
[Crustacea                 ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_abdA               ??PNGCPRRRGRQTYTRYQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINEQARKDR
Elminius_abdA              ????????????????????????????????????????????????????????????????????????????????
Porcellio_abdA             AGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQVKIWFQNRRMKLKKELR????????????????
Sacculina_abdA             ????????????????????????????????????????????????????????????????????????????????
Trypetesa_abdA             ????????????????????????????????????????????????????????????????????????????????
Ulophysema_abdA            ??????????????????????????????HFNHYLTRRRRIEIAHALCLTERQIKI????????????????????????????
[Chelicerata               ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_abdA          ????????????????????????????????????????????????????????????????????????????????
Cupiennius_abdA            AGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKEMRAVKEINEQARMES
Archaearenaria_abdA        ??????????????????????????????HFNHYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Steatoda_abdA              ??????????????????????????????HFNHYLTRRRRIEIAHALCLTERQIKI????????????????????????????
[Myriapoda                 ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_abdA          ?????CPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKEMRAVKEINEQARREA
Lithobius_abdA             ????????????????TRFQTLELEKVFHYNRYLTRRRRIEIAHTLVLSERQIKIWFQNRRMKLKKEMRAVKEINEQARREA
Pachymerium_abdA           ????????????????TRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIW????????????????????????????
Pauropus_abdA              ??????????????????????????????FHFNHYLTRRRRIEIAHALCLTERQIKI????????????????????????????
Scutigerella_abdA          ??????????????????????????????FHFNHYLTRRRRIEIAHALCLTERQIKI????????????????????????????
[Nematoda                  ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_egl_5              ????????????????????SVLEKEYQQSSYVSKKQREELRLQTNLTDRQIKI????????????????????????????
```


Brugia_egl_5 HWPNYAQSKKGRQTYQRYQTSVLESKFQQSSYVKTYQREELRLQTNLSDRQIKIWFQNRRMKAKKEKNRCEN?????????
Caenorhabditis_egl_5 AWPNYASSKKGRQTYQRYQTSVLEAKFQQSSYVSKKQREELRLQTQLTDRQIKIWFQNRRMKAKKEKQRVDDHTEHTPLLP
Meloidogyne_abdA ??
Pristionchus_egl_5 ??????????????????QTSVLEQKFLQSSYVSKKQREELRMQTNLTDRQIKI????????????????????????????
Strongyloides_egl_5 ???????????????RYQTSVLESKFQQSSYVSKKQREELRLQTNLTDRQIKI?????????????????????????????
Trichinella_egl_5 ???????????????ISTSVLESKFQQSSYVSKKQREELRLQTNLSDRQIKI?????????????????????????????
[Tardigrada ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Isohypsibius_abdA ????????????????????EFEKEFQFNHYLTRKRRIEIAHSLCLTERQIKI????????????????????????
[Onychophora ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Acanthokara_abdA ??TNGCPRRRGRQTYTRYQTLELEKEFHFNHYLTRRRRIEIAHVLCLTERQIKIWFQNRRMKLKKELRAVKEIN???????
[Nematomorpha ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Paragordius_abdA ??????????????????YQTLELEKEFRYNHYLTRRRRLEISHNLCLTERQIKI?????????????????????????????
[Priapulida ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Priapulus_abdA ??
[Lophotrochozoa ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Lineus_Lox4 ??
Girardia_DthoxF ???NSVQKRRGRQTYSRHQTLELEKEFQFNHYLTRRRRIEIAHNLCLSERQIKIWFQNRRMKLKKERQQ???????????????
Lingula_Lox4 ???????RRRGRQTYSRFQTLELEKEFQFNHYLTRKRRIEVAHALCLTERQIKIWFQNRRMKLKKERQQIKEINDTLKKPI
Nereis_Lox4 ??????EEERGRQTYSRYQTLELEKEFRFNHYLTRKRRMEIAHVLCLTEHQIKIWFQNRR?????????????????????
Patella_Lox4 ??????????????????????????????QFNHYLTRKRRIEIAHTLCLTERQIKIWFQNRRMKMKRERQAIKDINGDT????
Hirudo_Lox4 ??PNSSQRRRGRQTYSRYQTLELEKEFQFNRYLTRKRRIEIAHCLCLTERQIKIWFQNRRMKVKKEKQQIKELNEVGKG??
[Deuterostomia ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Mus_no ??
Branchiostoma_Hox8 ?????PERRRGRQTYSRYQTLELEKEFHFNKYLTRRRRIEIAHALGLTERQIKIWFQNRRMKLKKEAAMLC??????????
[10 20 30 40 50 60 70 80]
[....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
[abdominal B]
[Hexapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_AbdB ??
Anopheles_AbdB ??
Bombyx_AbdB DWTGQVTVRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLNLTERQVKIWFQNRRMKNKKNSQRQAAQAA??????
Drosophila_AbdB EWTGQVSVRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLQLTERQVKIWFQNRRMKNKKNSQRQANQQNNNNNSS
Folsomia_AbdB EWTGNVTVRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLNLTERQVKIWFQNRRMKNKKNSQRQQAL????????
Junonia_AbdB ??
Schistocerca_AbdB EWTGQVTVRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLNLTERQVKIWFQNRRMKNKKNSQRQAASSSSSNSSS
Thermobia_AbdB ????????RKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLNLTERQVKIWFQNRR?????????????????????
Tribolium_AbdB EWTGQVTVRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLNLTERQVKIWFQNRRMKNKKNSQRQAAQQQNNNNA?
[Crustacea ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_AbdB ??????TVRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLNLTERQVKI?????????????????????????????
Elminius_AbdB ????????????????????????????LFNAYVSKQKRWELARNLNLTERQVKI?????????????????????????????
Porcellio_AbdB ??
Sacculina_AbdB ???????VRKKRKPYSKYQTLELEKEFLFNAYVSKLKRWELARNLNLTERQVKI?????????????????????????????
Trypetesa_AbdB ??????????????????????????LFNAYVSKQKRWELARNLNLTERQVKI?????????????????????????????
Ulophysema_AbdB ??????????????????????????LFNAYVSKQKRWELARNLNLTERQVKI?????????????????????????????


[Chelicerata
Archegozetes_AbdB
Cupiennius_AbdB
Archaearenaria_AbdB
Steatoda_AbdB
[Myriapoda
Ethmostigmus_AbdB
Lithobius_AbdB
Pachymerium_AbdB
Pauropus_AbdB
Scutigerella_AbdB
[Nematoda
Ascaris_php_3
Brugia_php_3
Caenorhabditis_php_3
Meloidogyne_AbdB
Pristionchus_php_3
Strongyloides_php_3
Trichinella_php_3
[Tardigrada
Isohypsibius_AbdB
[Onychophora
Acanthokara_AbdB
[Nematomorpha
Paragordius_AbdB
[Priapulida
Priapulus_AbdB
[Lophotrochozoa
Lineus_Hox9
Girardia_post1
Lingula_post1
Nereis_post1
Patella_post1
[Deuterostomia
Mus_HoxA9
Branchiostoma_Hox9
[10 20 30 40 50 60 70 80]
[caudal]
[Hexapoda
Aedes_cd
Anopheles_cd
Bombyx_cd


Drosophila_cd QPGKTRTKDKYRVVYTDFQRLELEKEYCTSRYITIRRKSELAQTLSLSERQVKIWFQNRRAKERTSNKKGSDPNVMGVGVQ
Folsomia_cd ??
Junonia_cd ??
Schistocerca_cd ??
Thermobia_cd ??
Tribolium_cd AIGKTRTKDKYRVVYTDLQRIELEKEFTFVSYITIKRKSELAENLGLSERQIKIWFQNRRAKERKQNKKRIEEKSQIDNL?
[Crustacea ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_cd ??
Elminius_cd ??
Porcellio_cd ??
Sacculina_cd ????????????????????????????RFNNYITIKRKLELSRILGLTDRQVKIWFQNRRAKERKQKRRLEETSKRQLPGQ
Trypetesa_cd ??
Ulophysema_cd ??
[Chelicerata ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_cd ??
Cupiennius_cd ??
Archaearenaria_cd ??
Steatoda_cd ??
[Myriapoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_cd ??
Lithobius_cd ??
Pachymerium_cd ??
Pauropus_cd ??
Scutigerella_cd ??
[Nematoda ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_pal_1 NNVRVRTSEKYRMVYTDYQRLELEKEFRTTQFINSERKSQLSSELQLTERQIKI????????????????????????
Brugia_pal_1 ?????RTKDKYRVVYSDYQRLELEKEFRTSQFINSERKSQLSSELQLTERQIKIWFQNRLVFT?????????????????
Caenorhabditis_pal_1 NNVRVRTADKYRMVYSDYQRLELEKEFHTSPFITSDRKSQLSTMLSLTERQIKIWFQNRRAKDRRDKQKIR??????????
Meloidogyne_cd ??
Pristionchus_pal_1 ????????????????????LELEKEFHMNQFINADRKADLATKLNLTERQI????????????????????????????
Strongyloides_cd ??
Trichinella_pal_1 ????????????????????????ELEKEFRTSQFINSERKSQLLSELHLTERQI????????????????????????????
[Tardigrada ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Isohypsibius_cd SGKTRTKDKYRVVYSDHQRILEFEKEFFKNKYISVKRKAELARNLLLSERQVKI????????????????????????????
[Onychophora ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Acanthokara_cd ????????????????????????????HYSRYITIRRKSELAQALNLSERQVKI????????????????????????????
[Nematomorpha ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Paragordius_cd ??
[Priapulida ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Priapulus_cd ??
[Lophotrochozoa ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Lineus_cd ???????PDKYRVVYTDYQRLELEKEFHYSRYITMNRKAELAKSLDLTERQIKIWFQNRRAKERKINKKKDVMVKEP????
Girardia_cd ??
Lingula_cd ??
Nereis_cd ??
Patella_cd ??

```
[Deuterostomia          ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Mus_cd                  ????????????????????????????????????????????????????????????????????????????????
Branchiostoma_cd        ???????KDKYRVVYSDHQRLELEKEFYSNKYITIKRKVQLANELGLSERQVKIWFQNRRAKQRKMA??????????????

[                               10        20        30        40        50        60        70        80  ]
[                       ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]

[engrailed1]
[Hexapoda               ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Aedes_en                ????????????????????????????????????????????????????????????????????????????????
Anopheles_en            KKRADSEEKRPRTAFSNAQLQRLKNEFNENRYLTEKRRQTLSAELGLNEAQIKIWFQNKRAKIKKSSSEKNPLALQLMAQG
Bombyx_en               KKKAAPEEKRPRTAFSGAQLARLKHEFAENRYLTERRRQSLAAELGLAEAQIKIWFQNKRAKIKKASGQRNPLALQLMAQG
Drosophila_en           PKDKTNDEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKSTGSKNPLALQLMAQG
Folsomia_en             ????????????????????????????????????????????????????????????????????????????????
Junonia_en              QKKMNPEEKRPRTAFSAAQLARLKHEFAENRYLTERRRQALASELGLAEAQIKIWFQNKRAKIKKATGQRNPLAMQLMAQG
Schistocerca_en         KRNKKPEEKRPRTAFSGEQLARLKHEFTENRYLTERRRQELARELGLNEAQIKIWFQNKRAKIKKAS????????????????
Thermobia_en            RKEKKPEEKRPRTAFTSEQLARLKQEFQENRYLTEKRRQALARDLKLNESQIKIWFQNKRAKIKKASGQKNPLALQL????
Tribolium_en            QGAPTAEEKRPRTAFSGAQLARLKHEFAENRYLTERRRQQLSAELGLNEAQIKIWFQNKRAKIKKASGTKNPLALQLMAQ?
[Crustacea              ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Artemia_en              DKAITPDEKRPRTAFTAEQLSRLKHEFNENRYLTERRRQDLARELGLHENQIKIWFQNNRAKLKKSSGQKNPLALQLMAQG
Elminius_en             ????????????????FTTEQLERLKREFDENRYLSEKRRQDLAQELGLNESQIKIWFQNKRAKIKKSSGQKPVLASRL????
Porcellio_en            KKRDKKDEKRPRTAFTSDQLARLKKEFQENRYLTEKRRQDLARDLGLNESQIKIWFQNKRAKIKKQVSGSKNPLALQLM??
Sacculina_en            RVSRKTDEKRPRTAFSSEQLQRLASEFTDNRYLSEERRQRLARQLGLNESQIKIWFQNKRAKLKKTIPDKPSLAKKLMEQG
Trypetesa_en            ????????????????FTSEQLERLRQEFEDNRYLTEKRRQDLAQELGLNESQIKIWFQNKRAKIKKTGGSVRAGLAVHL???
Ulophysema_en           ????????????????????????????????????????????????????????????????????????????????
[Chelicerata            ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Archegozetes_en         ???????EKRPRTAFTAEQLARLKQEFQENRYLTEKRRQDLAKDLKLNESQIKIWFQNKRAKIKKSTGSRNPLAMHLMAQ?
Cupiennius_en           KKDKKPDDKRPRTAFTADQLSRLKHEFQENRYLTERRRQDLAKDLQLNESQIKIWFQNRRAKLKKASGQRSALALQLMAQG
Archaearenaria_en       KTSDGKADKRPRTAFSSEQLNRLRQEFSENRYLTERRRQDMARDLKLNESQIKI?????????????????????????????
Steatoda_en             ????????????????????????????????????????????????????????????????????????????????
[Myriapoda              ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ethmostigmus_en         ??????????????????????????QENRYLTEKRRQDLARELKLNESQIKI?????????????????????????????
Lithobius_en            ????????????????????????????????????KRRQDLARELKLNESQIKI?????????????????????????????
Pachymerium_en          ????????????????????????????????????????????????????????????????????????????????
Pauropus_en             ????????????????????????????????????????????????????????????????????????????????
Scutigerella_en         ????????????????????????????????????KRRQDLARELKLNESQIKI?????????????????????????????
[Nematoda               ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Ascaris_en              ????????????????????????????????????????????????????????????????????????????????
Brugia_en               ????????????????????????????????????????????????????????????????????????????????
Caenorhabditis_ceh_16   SSGSSEEEKRPRTAFTGDQLDRLKTEFRESRYLTEKRRQELAHELGLNESQIKIWFQNKRAKLKKSTSSVPRDRCSSVTPN
Meloidogyne_en          ????????????????????????????????????????????????????????????????????????????????
Pristionchus_en         ????????????????????????????????????????????????????????????????????????????????
Strongyloides_en        ????????????????????????????????????????????????????????????????????????????????
Trichinella_en          ????????????????????????????????????????????????????????????????????????????????
[Tardigrada             ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]
Isohypsibius_en         ????????????????????????????????????????????????????????????????????????????????
```


[Onychophora
Acanthokara_en
[Nematomorpha
Paragordius_en
[Priapulida
Priapulus_en
[Lophotrochozoa
Lineus_en
Girardia_en
Lingula_en
Nereis_en
Patella_en
[Deuterostomia
Mus_en
Branchiostoma_en
[odd hoxes left over]
Drosophila_ZEN2
Tribolium_ZEN2
Lithobius_PB2
Lithobius_SCR2
Thermobia_EN2
Elminius_ENb
Porcellio_ENb
Sacculina_ENb
Junonia_HOXa
Artemia_HOXa
Pachymerium_HOXa
Lithobius_ANTP2
Elminius_ANTP2
Ulophysema_ANTP2
Porcellio_UBX2
Cupiennius_UBX2
Porcellio_ABDA2
Porcellio_ABDA3
Acanthokara_ABDB2
[40]

```
Trypetesa_unplugged       ????????????????????????????????HSKKYLSLTERSQIAHALKLTEVQVKI?????????????????????????????
Lithobius_unplugged       ????????????????????????????????FHSKKYLSLTERSQIAHSLKLSEVQVKIWF?????????????????????????

Priapulus_HB3_Lox2_Ubx    ???????RRRGRQTYTRFQTLELEKEFHFNRYLTRRRRIEIANMLCLTERQIKIWFQNRRMKWKKEKQAYGATCSTETLT? [73]
Priapulus_HB2_Lox2_Ubx    ??????DRKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIANMLCLTERQIKIWFQNRRMKWKKECKARLIPNGDFPKEL [77]
Priapulus_HB1             ????????????????????????????FHFNRYLTRRRRIEIANVLRLTERQIKIWFQNRRMKWKKEHRLPKPVIPPPFPGD [58]
Priapulus_HB5             ????????????????????????????QFNQYITRARRVELARSLDLTERHLKI?????????????????????????? [27]
Priapulus_HB4             ARHRASFVRSKRKPYSKQQIVDLEKEYKVNPYITRQRRFEMAHNLGLSERQVKIWFQNRRMKTKKHLLTLRV????????? [76]

Lingula_Post_2            ???????????RKPYTRYQTMVLENEFLNNAYITRQKRWEISCKLHLSERQVKVWFQNRRMKRKKLNERAKALFKSEDIVV [71]
Nereis_post_2             ???????QRKKRKPYTRYQTMVLENEFMGNSYITRQKRWEISCKLHLSERQVKVWFQNRRMKRKKLNERAKTLIKS????? [69]

Lingula_antlike           ???????RKRGRHTYSRHQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKENKGIELSRDSPPQGV [75]

Branchiostoma_Hox_11      ?????KSTRKKRCPYTKYQTLELEKEFLFNMFVTRERRQEIARQLNLTDRQVKIWFQNRRMKMKRMKQRAM?????????? [66]
Branchiostoma_Hox_14      ?????KPVRPKRRPYSKYQLNELENEYVQNQYISRDKRLQLSQKLNLTERQVKIWFQNRRIKQKKLD????????????? [62]
Branchiostoma_Hox_13      ?????RGGRKKRCPYTKYQLSVLEQEYIQNRYVSRETRLELSQRLNLTDRQVKIWFQNRRMKQKRLE????????????? [62]
Branchiostoma_Xlox        ???????NKRTRTAYTRGQLLELEKEFHFNKYISRPRRIELAAMLNLTERHIKIWFQNRRMKWKKEQ????????????? [60]
Branchiostoma_Gsx         ???????SRRMRTAFSSTQLLELEREFASNMYLSRLRRIEIATFLNLSEKQVKIWFQNRRVKHKKEA????????????? [60]
Branchiostoma_Hox_10      ?????RVGRKKRCPYTKYQILELEKEFLFNMYVSRERRQEISRHVNLSDRQVKIWFQNRRMKMKRMN????????????? [62]

Mus_HoxA10                NWLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEISRSVHLTDRQVKIWFQNRRMKLKKMNRENRIRELTANFNF
Mus_HoxB11                GGSGGQRTRKKRCPYTKYQIRELEREFFFSVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKINRDRLQYYSANPLL?
Mus_HoxD12                GGAAPGRARKKRKPYTKQQIAELENEFLVNEFINRQKRKELSNRLNLSDQQVKIWFQNRRMKKKRVVQREQALALY?????
Mus_HoxA13                DASSYRRGRKKRVPYTKVQLKELEREYATNKFITKDKRRRISATTNLSERQVTIWFQNRRVKEKKVIKLKTTS????????

Caenorhabditis_nob_1      PWAISHDGKKKRQPYKKDQISRLEYEYSVNQYLTNKRRSELSAQLMLDEKQVKVWFQVCYRQYSAESLEIINNRRMKDKKL [81]

Meloidogyne_LIN_392        ??????????????????????ELEKEFHFNKYLTRKRRIEIAHTLILTERQVKIWF????????????????????????????

Brugia_cad_2              ??????????????????????ALELEKEFFRTKYISVKRKAEMARTLQLSERQVKI???????????????????????????????? [39]

Trichinella_ant_5         ??????????????????????ALELEKEFHYIKYLTLPRRIELTRTLGLTD?????????????????????????????????

[                                 10        20        30        40        50        60        70        80  ]
[                         ....'....|....'....|....'....|....'....|....'....|....'....|....'....|....'....|..]


;
endblock
;
```

Table 1 Species names of abbreviations used in the alignment of Hox genes and in phylogenetic trees

Hexapoda	
Aedes	*Aedes aegypti*
Anopheles	*Anopheles gambiae*
Bombyx	*Bombyx mori*
Drosophila	*Drosophila melanogaster*
Folsomia	*Folsomia candida*
Junonia	*Junonia coenia*
Schistocerca	*Scistocerca gregaria*
Thermobia	*Thermobia domestica*
Tribolium	*Tribolium castenatum*
Crustacea	
Artemia	*Artemia salina*
Elminius	*Elminius modestus*
Porcellio	*Porcellio scaber*
Sacculina	*Sacculina carcini*
Trypetesa	*Trypetesa lampas*
Ulophysema	*Ulophysema oeresundens*
Chelicerata	
Archegozetes	*Archegozetes longisetosus*
Cupiennius	*Cupiennius salei*
Steatoda	*Steatoda triangulosa*
Myriapoda	
Ethmostigmus	*Ethmostigmus rubripes*
Lithobius	*Lithobius atkinsoni*
Pachymerium	*Pachymerium ferrugineum*
Pauropus	*Pauropus sp*
Scutigerella	*Scutigerella immaculata*
Nematoda	
Ascaris	*Ascaris suum*
Brugia	*Brugia malayi*
Caenorhabditis	*Caenorhabditis elegans*
Meloidogyne	*Melodogyne javanica*
Pristionchus	*Pristionchus pacificus*
Strongyloides	*Strongyloides ratti*
Trichinella	*Trichinella spiralis*
Tardigrada	
Isohypsibius	*Isohypsibius ED*

Onychophora Acanthokara	
Nematomorpha Paragordius	*Paragordius robusta*
Priapulida Priapulus	*Priapulus caudatus*
Lophotrochozoa Lineus Girardia Lingula Nereis Patella	*Lineus sanguineus* *Girardia tigrina* *Lingula anatine* *Nereis virens* *Patella vulgata*
Deuterostomia Mus Branchiostoma	*Mus musculus* *Branchiostoma floridae*

Gene sequences were extracted from Genbank by searching with species and phylum names and looking for Hox genes in the submitted sequences.

APPENDIX 4 RAW DATA OF MICROFILARIAL RELEASE COUNTS OF CONTROL VS SHP-1 DSRNA TREATED BRUGIA MALAYI ADULT FEMALES

Worms (2 control worms and 2 treated with shp-1 dsRNA) were cultured in 10 ml of RPMI for 3 hours at the end of a 48 hour period following an initial 24 hours in dialysis culture (as described in chapter 5). The worms were then removed and the medium spun down to pellet the Mf. The Mf were resuspended in 200 µl of RPMI and 6 samples of 15 µl each were taken from both control and treated samples and the number of microfilaria counted under a binocular microscope. The remaining Microfilaria were placed on a slide (70 µl), air dried, fixed in methanol and Giemsa stained (as described in chapter 5) and counted under a microscope. During these counts the phenotypes of released Mf were also assessed. Averages were rounded to the nearest whole number (see table 5.3, chapter 5)

	Control worms	shp-1 dsRNA treated worms
15 µl sample 1	19	8
15 µl sample 2	21	11
15 µl sample 3	18	16
15 µl sample 4	23	12
15 µl sample 5	24	15
15 µl sample 6	15	10
Mf released per hour/ worm	(19+21+18+23+24+15) x 200/ (3 x 2 x 130) = 31	(8+11+16+12+15+10) x 200/(3 x 2 x130) = 17
Number of Mf on fixed slide (70 µl)	65	40
Number of short Mf on slide (70 µl)	4	19
Number of normal Mf released per hour/ worm	61 x 200/ (3 x 2 x 70) = 29	21 x 200/ (3 x 2 x70) = 10
Number of short Mf released per hour/ worm	4 x 200/ (3 x 2 x 70) = 1.9	19 x 200/ (3 x 2 x70) = 9